Three Roads to Objective Probability[1]

Preamble: Subjective Probability

The probability calculus can be put to many uses. One of these is for the representation of states of belief. There is an obvious sense in which our states of belief differ with respect to how confident we are in a proposition, and that confidence changes upon receipt of information. A physicist, for example, may regard a particular theory of dark matter as more or less plausible, and will adjust that assessment on the basis of both theoretical and experimental results. Confirmation theory is a normative account of how our confidence ought to change in such circumstances. It is convenient to describe such principles belief dynamics mathematically, which requires a mathematical representation of "degree of belief". In the usual idealization, complete and unshakable belief is represented by the number 1, total and utter disbelief by the number 0, and various intermediate grades of belief by real numbers between 1 and 0. A rule for adjusting strength of belief can then be represented by an equation, e.g. Bayes' rule.

Already we should note that some contentious assumptions have been made. Perhaps there just are no such psychological states as "total and utter disbelief" and "complete and unshakable belief". As a purely descriptive matter, there seem to be no psychological states that match the description: people can be brought to doubt, and to grant some credence to, anything. According to some *normative* theories, though, such states are theoretically possible, and would be impossible to *rationally* dislodge. If one endorses only Bayesian conditionalization as a means to changing states of belief, then a proposition accorded the strength of belief designated by zero or one would have to retain that strength forever. The Bayesian could, of course, admit these extreme degrees of belief as idealizations, adding that they never could, or should, actually obtain.

Much more contentious, and indeed frankly incredible, is the idea that the different psychological states can be correctly characterized by *unique* real numbers between zero and one. I believe in evolution much more strongly than in string theory, so in this representation a higher number ought to be attached to the former than the latter.

---

[1] Parts of this paper appeared as Maudlin (2007).

But there are certainly no *particular* real numbers that correctly characterize my psychological state. The usual response to this problem is to try to operationalize the assignment of numbers: ask me, for example, how much I would pay for a ticket that pays off $100 if and only if string theory is true. But I have no psychological disposition to assign a real number monetary value for such a ticket. I have no idea how I would react to such an offer in real life, and absolutely no reason to believe that my reaction would have any stability against varying details of circumstance that have no bearing on how strongly I believe the theory. So what we have here is evidently an idealization for pure mathematical convenience: by pretending that degrees of confidence can be uniquely characterized by real numbers, we make available various convenient mathematical tools for representing the dynamics of belief. First among these tools is the probability calculus. Bayes' rule again provides a good example: having accepted the fiction that my degrees of confidence correspond to real numbers, Bayes' rule can be used to specify a normative theory of how my confidence ought to behave upon receipt of new information. The rule gives good qualitative results: for example, finding out that a prediction of the theory turned out true ought to raise one's confidence in the theory.

I have provided this rough sketch of how the probability calculus gets imported into psychological theorizing in order to have a clear foil to other uses of the probability calculus, such as those that occur in physics. The use of the calculus in either descriptive or normative cognitive theory is, in a straightforward sense, subjective: it concerns the actual or ideal subjective psychological attitudes of cognizers toward propositions. "The probability assigned to string theory" is, in this locution, a way of talking about some person's actual, or recommended, *attitude* towards string theory. In a world without cognizers, without psychological states, without a notion of "available evidence", there would be no such thing as "the probability assigned to string theory". Such a world would, of course, either be a stringy world or a non-stringy world: string theory would be either true or false in it. But there would just be no sense in asking how *likely* string theory is to be true in such a world, or what its *probability* is. The probabilities, in this usage, are attached to propositions only as a means of representing psychological states: in the absence of psychological states they have no significance at all.

Still, a world without cognizers could have probabilities in it. In particular, there could be probabilities that arise from fundamental physics, probabilities that attach to actual or possible physical events in virtue solely of their physical description and independent of the existence of cognizers. These are what I mean by *objective probabilities*. Candidates for objective probabilities come in two classes: dynamical chances and reductive chances. The former are rather easy to describe and the latter more problematic. We will investigate them in turn.

Objective Probability 1: Stochastic Dynamics

Systems evolving in time are governed by a dynamics: laws concerning how the state changes with time. In a deterministic system, specification of the state of the system at one time together with the dynamics determines the state at later times.[2] In an indeterministic system, the state of a system at one time and the laws are jointly compatible with different states at later times. In a stochastic dynamics, these various possible evolutions of the system are assigned different probabilities.

Stochastic dynamics has been most strongly associated, in the history of physics, with quantum theory ("God plays dice."). And many of the ways of understanding that theory (although not all) do employ a stochastic dynamics. It is often said, for example, that there is absolutely no preceding cause that determines exactly when a radioactive atom will decay: two such (isolated) atoms could be *perfectly identical in all physical respects* at one time yet decay at different times in the future. If so, then their dynamics is indeterministic.

There is, however, still a probabilistic law that is supposed to govern these decays. In particular, there is supposed to be a fixed probability density (i.e., a fixed probability per unit time) that an atom will decay. Such a fixed probability density leads to the usual exponential decay formula, and underwrites the assignment of a *half-life* to each species of radioactive atom. The half-life is the period of time, from a given starting point, it takes for the atom to have had a 50% chance of decay.

---

[2] There are all sorts of subtleties here that won't much concern us. The *locus classicus* for a discussion is John Earman's *A Primer on Determinism* (1986).

Before we undertake a philosophical investigation of this sort of probability, we should note how different the use of the probability calculus is here from its use in the psychological theorizing discussed above. In the psychological case, it is frankly incredible that there is a unique real number that correctly characterizes my "degree of belief" in string theory. But in the radioactive particle case, it is directly posited that there is a unique real number that characterizes the probability density for decay, a number that in turn determines the half-life. Indeed, it is part of experimental practice in physics to determine half-lives to every greater degrees of accuracy. One might quibble about whether there is really a unique real number that can be assigned to the probability density, but one cannot deny that half-lives are determined, experimentally, to within parts-per-ten-thousand. The half-life of tritium, for example, is about 4499 days[3]. Further experimentation could refine the number by some orders of magnitude, if not indefinitely. Scientific practice proceeds as if there is a real, objective, physical probability density here, not just a mathematical model that has been imported to represent a looser structure, as in the case of stronger and weaker degrees of belief. The value of the half-life of tritium has nothing to do with the existence or otherwise of cognizers.

The most straightforward account of this probability density appeals to fundamentally stochastic dynamical laws. The behavior of tritium is a consequence of the laws governing its sub-atomic constituents, but that complication need not detain us: if the fundamental laws are irreducibly stochastic, then they can either directly assign or indirectly entail a probability density for decay, and hence allow one to calculate the likelihood of decay for any tritium atom over a specified period of time. The laws would allow initially identical tritium atoms to decay at different times, assigning a probability to a decay within any specified period. The probabilities involved, as they are *transition chances*, are *conditional* probabilities: they specify how likely it is that a system will evolve in a certain way *given that it started in a particular physical state*. In contrast, the "probabilities" used in the psychological theory are unconditional: they simply characterize strength of belief at a certain time.

At this point in the narrative, the usual philosophical move would be to allow that the notion of irreducibly stochastic dynamical laws is clear enough for the practical uses

[3] See Lucas and Unterweger (2000).

of physicists, but still involves some deep metaphysical questions, or confusions, or unclarities, which it the job of the philosopher to articulate and resolve. Unfortunately, I hold no such brief. I think that the notion of irreducibly stochastic dynamical laws, as postulated by physicists, is perfectly clear and requires no philosophical elucidation at all. It is rather the other way around: the evident coherence and clarity of the notion of fundamental transition chances can be used to help diagnose philosophical confusions.

What does the physicist do with a postulated stochastic dynamics? The game, once the dynamics is postulated, is to assign probabilities to various possible physical histories. In the case of the single tritium atom, this means assigning probabilities to decays in any specified interval of time, which we know how to do. Any individual tritium atom might decay before, or after, the half-life, and the probabilities for any specified result are calculable. If we have a pair of tritium atoms, they are treated as independent, so the joint probability distribution for the pair of decays is easy to calculate. Again, they might both decay earlier than the half-life, or both later, or one earlier and one later, with calculable probabilities for each. And it is no more difficult to calculate the probabilities for any number of tritium atoms, *including the total number of all tritium atoms in the history of the universe*. Given a finite number of such atoms, there will be a definite calculable finite probability that *all* the tritium atoms there ever have been or ever will be will decay well before the half-life of tritium. This is a straightforward *physical possibility* according to this particular stochastic dynamics.

At this point that the philosopher is likely to become uneasy. If *all* the tritium atoms decay within a certain time, relative to what standard are we entitled to call all of the decays "early"? After all, experimenters investigating the half-life of tritium would never conclude that all the decays were early: they would conclude that the half-life of tritium is much less than the number derived from the stochastic law. They would, that is, conclude that the atoms are governed by a different stochastic dynamics than the one we used to derive the probability. Doesn't that show that something has gone wrong?

But the philosopher has been misled here. All the atoms might have decayed early, where "early" is defined relative to the half-life, and the half-life is derived from the fundamental probability density. It is true that experimenters who lived in such a world would come to incorrect conclusions about what the fundamental probability density is. They would have been misled by a cosmic run of bad luck. That would be an epistemically unfortunate event for them, given that they wanted to arrive at the true probability density. But there is no insurance policy against bad epistemic luck. If the world came into existence just a few years ago, in the appropriate state (apparent fossils in place, etc.), we will never know it: bad luck for us.

The physicists in a "bad luck" world will be misled because in all *non*-probabilistic respects their world is exactly like other worlds with different fundamental probability densities. Indeed, any particular distribution of tritium decays is compatible with any fundamental probability density for the decay, although the different probabilistic laws will assign vastly different probabilities to the distribution. If one insists that only the non-probabilistic features of a world are to be included in the "Humean mosaic", then we have a clear breakdown of Humean supervenience: two worlds could agree completely in their Humean mosaic but disagree in their physical structure, since they are governed by different probabilistic laws. Since the evidence available to scientists in a world *is* determined by the Humean mosaic, at least one set of scientists will be rationally led to false conclusions about their world.

This failure of Humean supervenience is anathema to many philosophers, but I can't see any good reason require it. Hume had a reason: he thought that all ideas were composed out of simple ideas, and all simple ideas were copies of simple impressions. Two worlds with the same Humean mosaic would produce the same set of impressions in the human inhabitants, and so the inhabitants *could not have* contentful thoughts that were true of one world but not the other. But nobody follows Hume's lead on the origin of ideas any more, so that is no grounds to insist on Humean supervenience.

Some of the logical empiricists arrived at a weaker cousin of Hume's semantics: the content of a statement was supposed to be specified by the means we have to confirm or disconfirm it. If no observations could provide evidence for or against a statement, then the statement was at the least non-scientific, if not flatly meaningless. Again, this semantic doctrine has taken a rather bad beating in the last half-century, but it is worthwhile to point out that fundamentally stochastic dynamics pass this logical empiricist test: we do know how to produce evidence that confirms or disconfirms hypotheses about stochastic dynamics. That is just what the scientists who try to determine the half-life of tritium are doing.

Indeed, it is precisely here, in the *testing* of probabilistic hypotheses, that the probability calculus, and Bayes' theorem, come into their own. The experimenter comes to the problem firmly convinced—never mind how—that there is a constant probability density for the decay of tritium, but not knowing what that density is. As we have already seen, every possible pattern of decay events in the laboratory will be *consistent* with every possible value for that density, and each density will assign a probability to each pattern of decays. But the probabilities will be wildly different: a probability density that yields a half-life near to the *observed* half-life of the tritium atoms will assign a much, much higher probability to that pattern than a probability density whose half-life is far from the observed half-life. And Bayes' theorem tells us what to do with that observation: using the pattern of actual decays as evidence and the various values for the probability density as the hypotheses, we derive the probability of the evidence on the various hypotheses. And we can then use Bayes' theorem and Bayesian conditionalization to tell us how to update our confidence in the various hypotheses.

The experimenter comes to the question in a particular psychological state. He does not know the exact half-life of tritium, but surely has some opinions: familiarity with tritium indicates a half-life on the order of twelve years. Half-lives significantly shorter or longer than this are ignored—the experimental design would not be appropriate for them. So the experimenter comes with a (vaguely defined) range of values within which the half-life is supposed to fall. His attitude to the values in the range will be about the same, with credibility dropping off as the values approach the (vaguely defined) boundaries. He will certainly *not* have a psychological state that corresponds, say, to a precisely defined probability density over the various half-lives. Still there will be many probability densities that could be used—with equal validity—to represent this psychological state: any density with most of the probability smoothly distributed over the target range and falling off outside it. Arbitrarily pick such a probability density to represent the psychological state. As the data comes in, Bayes' rule can be used to update this probability denstity, and the usual merger theorems give us the right qualitative behavior: after enough data has been conditionalized on, such representation of the initial psychological state will become very sharply peaked around values for the probability density that yield half-lives close to the observed half-life. The posterior probability for values far from the observed half-life will be driven quite low, but never to zero. This represents the psychological state of someone who very strongly believes the true half-life to fall in a small interval, but who cannot absolutely rule out the proposition that it falls arbitrarily away. This is exactly the attitude we endorse.

So despite the failure of Humean supervenience (in this sense), despite the fact that different probabilistic hypotheses are compatible with every set of evidence, there is a straightforward account of how probabilistic hypotheses are tested, confirmed, and disconfirmed. It is a consequence of these methods that no hypothesis will ever be absolutely ruled out, no matter how much evidence is collected. Even a perfectly rational omnisentient being, who observed every observable event from the beginning of time to the end of time, would not know for certain what the half-life of tritium is. But such a being would have a very strong opinion on the matter, and would be fully justified in having it.

This account of stochastic dynamics has not offered, and will not offer, any *reductive analysis* of objective physical probabilities to anything else. It cannot, since worlds with different objective physical probabilities can display the very same non-probabilistic set of facts. But neither, I think, need it do so. If someone complains that he doesn't understand the notion of physical probability here, then I am tempted to respond with Dr. Johnson: "I have found you an argument, I am not obliged to find you an understanding". That is, I cannot deny the possibility of a sort of cognitive blindness that would make someone unable to comprehend the notion of probability being used here, and I cannot offer a remedy for such blindness, since the notion here appears as an irreducible posit. But such cognitive blindness appears to be rare: when the notion of a stochastic dynamics is introduced to the uninitiated, the result is not blind incomprehension. Some, like Einstein, might not *like* the idea, but they *understand* it. Furthermore, it is not at all clear what is wanted to provide the needed clarification. It is clear how hypotheses about stochastic dynamics are to be formulated, used, and tested. It is clear how to do experiments and draw conclusions. No reductive analysis is offered because the notion is not a derivative one, and there have to be non-derivative concepts. What more, exactly, could be wanted?

Objective probability in the sense of stochastic dynamics is evidently what David Lewis meant by the term "objective chance". In Postscript B to "A Subjectivist's Guide to Objective Chance", Lewis addresses a problem raised by Isaac Levi, "which problem, it seems, is the reconciliation of chances with determinism, or chances with different chances" (Lewis 1986 p. 117). Lewis makes short work of the reconciliation project: "To the question of how chance can be reconciled with determinism, or to the question of how disparate chances can be reconciled with one another, my answer is: *it can't be done*" (p. 118). This indicates that Lewis thought of chance purely in terms of stochastic dynamics. We will take up the problem of "deterministic chances" below.

Objective Probability 2: Humeanism

Since stochastic dynamics, as I have elucidated it, is incompatible with Humeanism, Humeans need some other account of objective probabilities. It goes something like this.[4]

The foundation of contemporary Humeanism is the conviction that there is some Humean base, often called the "Humean mosaic", of non-nomic, non-probabilistic fact, upon which all other facts supervene. There is a certain amount of haggling that can go on about the exact credentials needed to get into the Humean base. Lewis thought that the basic Humean facts were also local facts—intrinsic features of pointlike or smallish regions or things—related only by spatio-temporal relations (thus a "mosaic"). That feature has been largely abandoned under pressure from modern physics: quantum theory requires a wavefunction, and the wavefunction is not local in this sense. Often the Humean base is characterized as "categorical" fact, as opposed to "dispositional", but these labels are really just placeholders that stand in need of clarification. If one believes in non-Humean laws (as I do), is their existence a categorical fact? I don't see how the meaning of "categorical", insofar as it has a clear meaning, rules this out. So one gets a specific form of Humeanism by articulating specific restrictions on the Humean base. It

[4] See Barry Loewer (2001) and (2004) for a much more detailed account.

appears to be required, at least, that the basic Humean facts not be intrinsically nomic or probabilistic. This ought to be taken to be just constitutive of contemporary Humeanism, since an epistemic or semantic principle that would admit wavefunctions and rebuff laws is not evident. Neither of these items is directly observable, and so both would have been rejected as fundamental by Hume.

Humeans don't want nomic facts in the Humean base, but neither do they want to eschew the claim that there are laws of nature. Propositions achieve the status of laws rather by playing a certain role in characterizing the Humean base. A complete account of the base, encompassing the entire history of the universe in all submicroscopic detail, is beyond our capacity, but still quite a lot of information about it can be conveyed very compactly. Laws of nature, on this account, are nothing other than those propositions that simultaneously maximize informativeness about the Humean base and minimize complexity in formulation (in a preferred language). For example, if Newton's Laws of mechanics were true, they, together with a specification of all force laws, would allow one to recover the totality of the Humean mosaic merely from some initial conditions. This would be highly informative.

The clearest cases for Humean information compression are provided by perfect correlations, but statistical regularities also could play the role. If one flips a coin a million times, it will (probably!) take a tremendous amount of description to convey the exact sequence of results. But if the results look statistically typical for a set of independent trials on a coin of fixed probability, then just specifying the probability will give quite a lot of guidance for what to expect. This is not frequentism: there is no requirement that the specified probability exactly match the frequency of outcomes, since some probability near the exact frequency may be easier to specify. Of course, the specified probability cannot diverge too far from the actual frequency, otherwise the expectations produced will lead one badly astray. And the category "flips of this coin" or even "flips of a coin" is not likely to be simply translatable into the language of fundamental physics, so the example is not realistic in any case. But the moral still stands: probabilistic language can be used to formulate compact, informative descriptions of ensembles, and so could properly be used in formulating Humean "laws".

It might appear that the adoption of Humeanism instead of the postulation of irreducible stochastic dynamics would make little difference to the practice of physics. Even if one believes that the world is governed by an irreducibly stochastic dynamics, the only empirical basis for determining the dynamics is observed frequencies. Bayesian conditionalization on observed frequencies, as we have seen, will concentrate one's strength of belief on fundamental probabilities that are close to the observed frequencies. So the difference in beliefs between a Humean and a non-Humean are likely to be rather subtle. Let's pretend, for the sake of argument, that the decay of tritium were the sort of event that could be covered by a specific law. (This is not realistic, as noted above, because tritium is composite, and the fundamental laws, according to both Humean and non-Humean, govern the components.) And suppose, as is not possible, that both the Humean and non-Humean knew the exact lifetimes of all tritium atoms in the history of the universe. The Humean and the non-Humean would very nearly agree about the half-life of tritium. The non-Humean would assign a very high credibility to the claim that the half-life is in a small interval centered on the *observed* half-life. If one of the values in that interval had a salient theoretical virtue (related, for example, to symmetries in the situation), that particular value might be deemed especially credible. The non-Humean would reserve some degree of credibility for the claim that the actual half-life is far from the observed value, but the degree of credibility would be so small as to make no practical difference. The Humean, on the other hand, would be able to make no sense of "the actual half-life": the ensemble of particular lifetimes of particular tritium atoms is all there is. Still, if the distribution of that ensemble has approximately a decreasing exponential form, the simplest way to convey information about it might be by means of a "probabilistic law". And considerations of simplicity or symmetry might militate in favor of a particular value for the probability density, a value whose derived half-life might differ, a bit, from the time it took half of the tritium atoms to decay. The Humean could presumably make no sense of spreading around credibility to different values for the half-life in this situation, since there is no fact of which he would be ignorant.

In real life, though, we never have complete information about, for example, the lifetimes of all tritium atoms. So the Humean must face the problem of extrapolating from observed instances to unobserved ones, reflecting incomplete information about the

Humean base. A Humean considering probabilistic laws for tritium decay on the basis of actual experiments will also have a spread of credibility among various candidates. And that spread might be quite extreme, reflecting the possibility that the observed decays have a distribution quite different from that of all decays. The psychological state of the real-life Humean and non-Humean, then, appear quite similar.

This examination of the psychological states of the Humean and the non-Humean does not render the probabilities at issue subjective. For the believer in stochastic dynamics, there is a mind-independent fact about what the dynamics is. For the Humean, there is a mind-independent fact about both what the Humean base is and about what the simplest, most informative ways of communicating its features are. In real life situations, both the Humean and non-Humean will always remain somewhat uncertain about these things, and their uncertainties will be similar in many respects.

Still, the differences between Humeanism and stochastic dynamics can show up in actual scientific practice. For the notion of stochastic dynamics is more specific, and more constrained, than the notion of a compact method for conveying information about the Humean base. Let's consider the differences.

As noted above, the probabilities that arise from a stochastic dynamics are all conditional probabilities: they reflect transition chances between an initial and a final state. Hence a stochastic dynamics cannot, by itself, have any implications about what the initial state of the universe is, since that state is not produced by a transition from anything else. But for the Humean, probabilistic claims are just a means of conveying information about empirical distributions. There is no reason to expect only transition probabilities to arise in service of this end: other probabilities might be useful as well. For example, about 1 in every 6000 atoms of hydrogen in the universe is deuterium. If they are uniformly spread about among the hydrogen atoms, then a simple way to convey something about the distribution of deuterium is to say that every hydrogen atom has a 1/6000 chance of being deuterium. This evidently has nothing to do with dynamical transitions, and so by itself could not be the basis of a stochastic dynamics. But it is a perfectly good candidate, as it is, for a Humean law. (The relative proportion of protium and deuterium is presumably a consequence of dynamics—of particle formation after the

Big Bang. So the Humean may not *need* this law, if there is already the appropriate Humean dynamical law. But that is a different matter.)

Even more importantly, there is nothing to prevent there being probabilistic Humean laws concerning the initial state of the universe. Since Humean laws are not intrinsically tied to dynamics, or to time, the initial state is just one more piece of the Humean basis to be summarized—a particularly critical one if there happen to be Humean dynamical laws. So the Humean could use a probability measure over possible initial states of the universe without any particular worries: it would just be a way to convey information about the distribution of Humean states of affairs there. A stochastic dynamics provides nothing comparable.

Humean probabilities have a wider area of application exactly because they carry no metaphysical baggage: all there is at bottom is the Humean mosaic. Some may judge that this greater Humean breadth is purchased at the price of the sort of ontological depth that underwrites explanation. It is exactly because an irreducible stochastic dynamics is ontologically independent of the non-probabilistic facts that it can be part of an account of the production of those facts. The observed distribution of lifetimes of tritium atoms is explained by an irreducible dynamics as a typical outcome of the operation of that dynamics. Humean laws, in contrast, can summarize but not, in the same way, explain.

This is not the place to pursue this dispute, which lies at the heart of many objections to Humeanism. It is enough that we have seen how Humean probabilities can arise as objective features of the world, and how they differ, both conceptually and practically, from irreducible stochastic dynamics.

Objective Probabilities 3: Deterministic Chances

To Lewis's ears, the phrase "deterministic chance" would be a *contraditio in adjeto*. Let's start with an example that appears to require the idea.

Scenario 1) The dynamics of the world is deterministic Newtonian dynamics. Part of the physical description of the world is this: from the beginning of time, there has existed a planet. On the surface of the planet is a large collection of boxes. In each box, there an object with the physical structure of a fair die: cubical, of uniform density, with different numbers of pips on each face. Throughout all time, the boxes have lain undisturbed on the surface of the planet.

Question: from the description given above, does the physics have any implications about the distribution of upward-facing sides of the dice?

Intuitive answer: No. The distribution is whatever it was at the initial time. Nothing in the physical description given has any bearing on that distribution.

Scenario 2) As in scenario 1, save for the following addition. At time T, there is a severe earthquake on the planet. The boxes are all tumbled around. (A complete physical description of the earthquake could be provided.)

Question: Given the new description, does the physics have implications about the distribution of upward-facing sides of the dice after T?

Intuitive answer: Yes. Just as an irreducibly stochastic dynamics with equal probability for each of six outcomes would have probabilistic implications, so in this deterministic case, there are the same probabilistic implications: Each die has an equal chance of being on each of its six faces, with the probabilities independent. If the ensemble of boxes is large, there is a very high chance that about one sixth of the ensemble will show each face, and a low (but easily calculable) chance that they all show the same face.

The dynamics in these scenarios is, by stipulation, deterministic. Nonetheless, the probabilistic conclusion in scenario 2, if it can be vindicated, is somehow a consequence of the dynamics. In scenario 1, the dynamics is evidently irrelevant, since the state of the dice never changes, and the intuition is that no probabilistic conclusion can justifiably be drawn.

Furthermore, the probabilistic conclusion clearly depends on the dice all being of uniform composition. If they were all loaded dice, with a weight embedded in the center of the face that has one pip, then scenario 1 would be unchanged, with no implications at all about the distribution of the dice, but in scenario 2, one would no longer accord equal weight to all six faces, nor expect about one sixth of the ensemble to show each face. One would expect more than a sixth to show six pips, and the appropriate statistical distribution would be a matter of some very complex calculation, in which the laws of Newtonian mechanics would figure. It is an open question at this point whether those calculations ought to yield as sharp a probability distribution over the possibilities as one would get using a stochastic dynamics (from a *specific* initial state), but it does appear that a weighting toward dice with six pips facing up ought to be derivable from purely physical considerations.

One could, of course, stick by Lewis's assessment: if the dynamics is deterministic, then *ipso facto* there are no objective probabilities, even in scenario 2. This would, however, run contrary to the usual understanding of physics. It is commonly thought, for example, that phenomenological thermodynamics was "reduced", in a certain sense, to statistical mechanics. In the course of the "reduction", the deterministic thermodynamical laws are *replaced* by probabilistic assertions. So whereas according to the laws of phenomenological thermodynamics it would impossible for a glass of lukewarm water to spontaneously segregate itself into an ice cube surrounded by boiling water, according to statistical mechanics such an evolution is possible but highly unlikely. Similarly, the various gas laws, or laws of chemical combination, are transmuted into claims about the most likely evolutions of systems described *deterministically* at a more fundamental physical level and treated using the tools of statistical mechanics. And the tools of statistical mechanics, whatever they are, are not supposed to advert to, or make use of, subjective states of credence. So the probabilities they employ should be considered objective in some appropriate sense.

After all, it is a plain physical fact that ice cubes do not spontaneously form in lukewarm water, that the air in rooms never spontaneously congregates in one corner, that tossed uniform dice never show the same face a thousand times in a row, etc., even though all of these behaviors are compatible with the basic dynamical laws. The fact that physical systems do not act this way has nothing to do with subjective states of belief: lukewarm water did not spontaneously freeze long before there were any subjective states. This appears to be the sort of straightforward physical fact that calls for a physical explanation. Since all of the unrealized behaviors are acknowledged to be physically *possible* such an explanation could do no better than to show that they are, in some objective sense, nonetheless highly unlikely.

The benchmark work for this problem is that of Boltzmann on thermodynamics. An exact understanding of what Boltzmann accomplished is a very difficult matter. But I do think I now appreciate what his arguments were not, and what is being attempted in the "reduction" of thermodynamics to statistical mechanics. What insight I have is derived from lectures given by Detlef Dürr[5] and conversations with Dürr, Sheldon Goldstein and Nino Zanghi, and I would like to do what I can at least to forestall a set of misperceptions that seem to be very widespread. At least, I fell prey to these misperceptions until very recently.

The basic misunderstanding about the foundations of these statistical arguments arises from the following thought. A deterministic dynamics effects a mapping from the initial state of the universe to its state at any later time. This map can be used to project a probability measure over possible initial states to a probability measure over later states, and so over later dynamical evolutions. So if one wants to show that some sort of evolution in a deterministic theory is "likely" or "unlikely", the only route is to find some preferred probability measure over possible initial states. The questions are then 1) What is the preferred measure? 2) Why is it preferred? and, most deeply, 3) What does it even mean to attribute a probability measure to the set of initial states of the universe? Given that there only was one initial state, and the probability measure cannot be characterizing some mechanism that produces initial states, what would such a probability measure correspond to physically?

If one falls into this way of thinking, then question 1 seems not so difficult: the preferred measure is, say Lebesgue measure[6] over some ensemble of states, or, in the case of Bohmian mechanics, the measure over configurations given by the square-amplitude of the universal wavefunction. Question 2 is rather more difficult, but something can be attempted: for example, the measure might be picked out by being stationary, or equivariant, under the dynamics. Indeed, a measure might even be picked out be being the *unique* stationary measure, so the deterministic dynamics is uniquely associated with it. But this approach to question 2 makes an even greater mystery out of question 3: if what makes the measure special, or preferred, is the *dynamics*, why in the world should it be associated with the set of possible *initial states* of the universe? After

[5] At the 4th International Summerschool, *Philosophy, Probability and Physics*, University of Konstanz, Germany, August 2005

[6] By "Lebesgue measure", I mean the standard volume measure over phase space, which is Lebesgue measure relative to canonical coordinates.

all, the initial state of the universe is exactly the unique state in the history of the universe that was *not* produced by any dynamics! By what magic pre-established harmony would a physically significant measure over these states arise *from the dynamics*?

Here is an example of the deleterious, and mistaken, reasoning that arises from this approach. Take Bohmian mechanics as an example. Since the theory is deterministic, it is said, probabilistic or statistical predictions can arise from the theory only if one puts a probability measure over the possible initial states—possible particle configurations—that form, together with the initial universal wavefunction, the initial state of the universe. The "appropriate" probability measure is one that will return Born's rule as the right rule for making predictions for the outcomes of measurements. But then it seems on the one hand that the absolute square of the initial wavefunction must uniquely give the "right" measure, and, correspondingly, that there is no mystery at all about why the "right" predictions come out. After all, if you put Ψ-squared in at the beginning of time, it is not so surprising that you get Ψ-squared out at later times. Indeed, all one needs to do is show that the trajectories in Bohm's theory follow the flow of Ψ-squared—the flow of the "probability density"—and voila[7]. The whole proceeding seems very simple, but also something of a cheat: you only get out the right answer because you put it in by hand in the initial measure. And "justifying" that measure by appeal to its *dynamical* properties appears to be, as noted above, unprincipled.

But the reasoning sketched above *is not the reasoning being used to justify the statistical predictions at all*. The matter is much more subtle and difficult, but also much more satisfying.

To begin with proving the negative: if the strategy sketched above were all there is to justifying the statistical predictions of Bohmian mechanics, then the justification would go through no matter what the initial wavefunction of the universe is. In particular, it would go through even if the initial wavefunction were real, i.e. if the whole universe were in its ground state. But given a real wavefunction, and the Bohmian dynamics, *nothing would ever move*. The situation would be completely analogous to situation 1

[7] Matters are not, in fact, nearly so simple! (Recall that I am presenting a *mistaken* view.) To understand the formalism of quantum mechanics from the point of view of Bohmian mechanics, one must first understand how *subsystems* of the universe get assigned wavefunctions, and how predictions are made from these.

above: after any amount of time, the particles in the universe would be exactly where they started out in the in initial state. And in this situation, the *only* possible justification for statistical predictions at later times would have to be that the very same statistical conditions hold in the initial state, just as the only reason to have any expectations about the distribution of the dice in situation 1 must derive from the very same expectations about their distribution in the initial state. But, as we have seen, the case of coins in the boxes does depend, intuitively, on the boxes being shook. So we should look for a reconstruction of the reasoning that makes use of this fact. Similarly, one would expect that Boltzmann's "reduction" of thermodynamics depends on there being *motion* in the universe: in the absence of such motion, no justification of statistical predictions is evident.

So the derivation of deterministic chances is a more subtle matter than just pulling out of thin air a probability distribution over initial conditions. Following Dürr's lectures as best I can, let me sketch how the argument goes.

Consider a Galton board or quincunx, i.e., the familiar board with pins in it, down which balls execute a random walk by ricocheting off the pins (see Figure 1, or visit teacherlink.org/content/math/interactive/flash/quincunx/quincunx.html for a nice computer simulation). Balls from a hopper at the top are fed into the board, and the balls at the bottom form, with very high probability, something close to a Gaussian distribution.

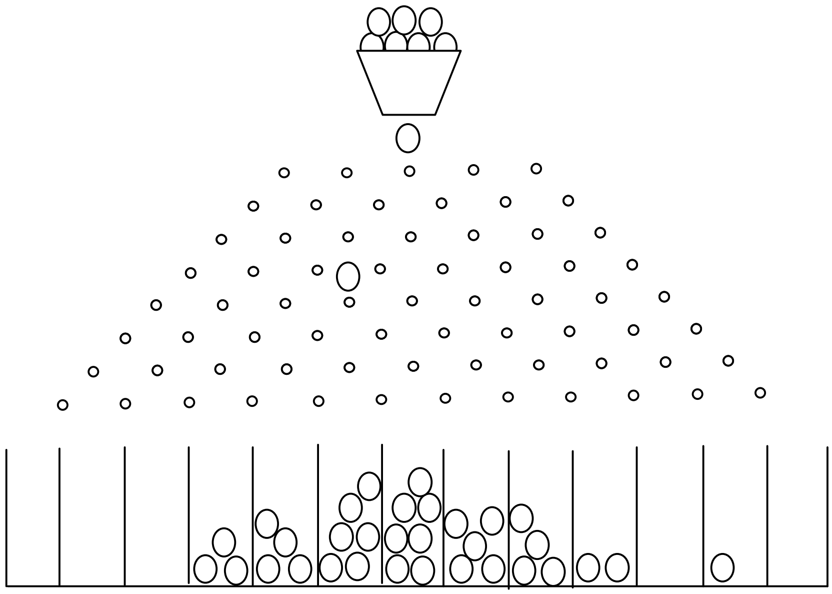

Figure 1: Galton Board

How can we analyze the Galton board mathematically?

First, the dynamics of the board is postulated to be deterministic. The balls collide elastically with the pins. The balls are idealized as perfect spheres, the pins as perfect cylinders, exactly evenly spaced. Evidently, these idealizations will not match any actual physical board, and the corrections needed to accurately represent a real board would make significant changes to predictions about the trajectory of any *particular* ball. Nonetheless, many corrections would not significantly affect the *statistical* predictions. Given the idealization, it is easy to get a sense for the dynamics. To begin with, we will consider just a single ball, released somewhere at the top. We will focus our attention on certain statistical features of the ball's trajectory as it makes its way down the board. At the most rudimentary level, we will consider just the proportion of times that the ball is deflected to the right and to the left when it reaches a pin. Evidently, the exact proportion—indeed the exact details of how the ball passes each particular pin—are determined by the exact initial position and velocity of the ball. But we want to try to derive conclusions that are somehow independent of that exact initial state. How can this be done?

For analytical purposes, let's suppress the variation in velocity and focus only on the variation in initial position. Since the balls come out in different places, we will begin by using a probability measure over initial positions to represent the hopper. This probability measure *is not* supposed to represent some physical fact in the situation: it is not as if, for example, a particular physical hopper is associated with some unique probability measure over the possible positions for balls starting at the top.

One might be tempted to suggest that there is some such unique measure: imagine running the hopper forever, feeding an infinite number of balls through it. Then there would be a limit, as the number of balls goes to infinity, of the proportion of balls whose starting position is in any given finite interval. This limit would give us the right probability measure over initial positions.

But a moment's thought reveals that this is just a fiction. There is no distribution that "there would have been had an infinite number of balls gone through". On the one hand, the hopper would break after a while, and on the other, the exact starting position of a ball would depend on many circumstances (wind, temperature, exact way the infinitude of balls had been stacked, etc.) for which there is no fact about "what they would have been". So we must recognize that treating the initial state of the ball by means of a probability distribution is not a matter of finding an exact mathematical representation of some precise physical condition. It is rather closer to the use of a probability distribution over propositions to represent a psychological state: the probability distribution is a *convenient* but *much too mathematically detailed* object to do the job we want. What we would like is for most of the precise mathematical details of the probability distribution we choose to wash out in the end, to be *irrelevant* to our final conclusions. We will see how this happens.

For the moment, though, let's just *choose* a convenient probability measure to represent the initial state of the ball. What does "convenient" mean here? Well, since we are concerned with *dynamics*, with calculating the measure of trajectories that have certain specified dynamical features, a convenient measure will be one that coheres in certain way with the dynamics. For example, the measure might be stationary under the dynamics, or might be *equivariant.* An equivariant measure has a fixed functional relationship to the state of a system at each time. We want a measure that will indicate, at any time, how "large" various regions in the phase space, or configuration space, of the system are. We need this because our conclusions are going to be about the features of a "large" number of microstates compatible with a given constraint. But then the notion of "large" should be defined at all times. And more importantly, we want the notion of "largeness" be such that *the dynamics will carry a large set at one time into a large set at another time*. This will evidently be true if the measure is stationary under the dynamics, but can equally well be achieved by using an equivariant measure. It is this feature that will allow us to draw conclusions about the size of the relevant sets at different times. In classical statistical mechanics, the measure used is Lebesgue measure on the phase space. As noted above, for our purposes we will focus just on position, and use a spatially flat measure there.

For our ball, it is not enough to specify a flat measure: we need an interval at the top, under the hopper, over which the measure is defined, roughly an area from the furthest to the right a ball will begin to the furthest to the left. And once again, there is clearly no fact about where the endpoint should "really" be. Even more, it is obvious that the flat measure will, in a sense, be *inappropriate* for the regions at the edge of the interval: balls will surely start out proportionally less often (per inch) at the extreme edges of the interval than at the center, so a flat distribution over the interval, however we choose it, is not realistic. But we still choose some interval—a convenient one—and use the flat distribution anyway. All will be well at the end.

Now what we would like to do is to see what happens to *most* of these possible initial states (with "most" defined by the measure) as the corresponding balls make their way down the board. Since the dynamics is very complex and chaotic, with each particular initial state leading to a complicated set of collisions with the pins, and initially nearby trajectories being carried far apart, this analysis cannot be carried out rigorously. And it will make our conclusions easier to state if we consider what would happen in the limit as the board becomes infinitely long. Then, by appeal to the symmetry of the pins and the nature of the dynamical interaction, we can argue with great plausibility for the following conclusion: if we imagine letting the Galton board become infinitely long, with an infinite number of rows of pins, then for *almost every* initial state (calculated by the flat measure over the chosen interval), the limiting frequency of deflections of the ball to the right when it hits a pin will equal the limiting frequency of deflections of the ball to the left when it hits a pin, both being 50%. Of course, there will be initial states for which this is not true: states in which, even in the limit, a ball is deflected more often to one side than to the other. There will initial states in which the ball is *always* deflected to the right, or *always* to the left. There will even be initial states in which there is no limiting frequency of deflections at all. But what is plausible—and what we in fact believe as a purely *mathematical* fact—is that the set of initial states with 50% limiting frequency for deflections to either side is a set of measure one. Let us grant, for the remainder of this discussion, that this mathematical claim is true. It will similarly be true that in the limit, successive interactions with the pins will be statistically independent. That is, if we look as what happens in two successive rows of pins, we will find, for a set of measure one, that the limiting frequencies of the sequences "Pass to the right then pass to the right", "Pass to the right then pass to the left", "Pass to the left then pass to the right" and "Pass to the left then pass to the left" will all be 25%. And so on.

Let's now introduce a piece of terminology. Let's say that when some specified dynamical behavior (like passing a single pin to the right, or passing successive pins first to the right and then to the left) has the same limiting frequency in a set of initial states that has measure one, that frequency for the dynamical behavior is *typical*. "Typical" is not a magical word here—it just characterizes a well-defined mathematical property of dynamical behavior in our model, relative to the initial probability measure. As we have seen, it would be a very, very, very difficult thing to *prove* typicality in the model we are discussing, but there are persuasive plausibility arguments. But right now, we are just introducing a word, by stipulation.

Nota Bene: the adjective "typical" is applied not to *individual initial states*, but to *behaviors* or *properties of trajectories*, such as have a given limiting frequency. The behavior is typical if it is displayed by the evolution of most of the initial states, with "most" understood relative to the measure. It is simply ungrammatical to say that a particular initial state is "typical": one can only say that it displays such-and-such a typical property. It is perfectly possible that almost *no* initial state displays *all* typical features, i.e. that it is typical for a system to have *some* atypical features.

The essential thing to see is that if we use "typical" in this way, *then the particular choice of flat measure at the outset, and the particular choice of the interval over which the flat measure was defined, become irrelevant*. *The very same frequencies would count as typical had we chosen any other measure over the interval, so long as it is absolutely continuous with the flat measure*. For absolutely continuous measures agree on which sets are sets of measure one and which are sets of measure zero. So had we allowed the measure to decrease toward the edges of the interval, reflecting the fact that for a real hopper proportionally fewer balls will start near the edges, the results would be exactly the same. It is also extremely plausible that the choice of the exact interval is irrelevant: make it larger or smaller, and still the same frequencies will be typical. In this case, our concerns about how to pick the "right" probability measure to represent the possible initial states of the ball, or even what the "right" measure *means*, very nearly evaporate: if you don't like the flat measure over the interval, pick any other absolutely continuous measure. If you don't like the exact endpoints of the interval, pick any other. All you have to do is avoid extremism: don't pick a new measure that concentrates finite probability on a set that got zero probability originally and don't shrink the interval down to a point.

There is no reason that typical behavior, as we have defined it, should have anything particularly to do with statistics. If we have two metal rods at different temperatures and then bring them into thermal contact, typical behavior will be for the motions of the atoms in the rods to evolve so that the temperatures in the rods equalize. This is the way that the laws of thermodynamics, which are *deterministic*, are "reduced" to statistical mechanics: thermodynamic behavior is shown to be typical behavior. In the case of the Galton board, though, the typical behavior is characterized statistically, so the behaviors involve proportions between 0 and 1. The proportions *are closely related* to the probabilities that are postulated in a stochastic dynamics. We must carefully consider what that relation is.

It seems, at first glance, that we have derived statements about typical behavior that have a similar abstract form as claims that arise from an irreducibly stochastic dynamics. Where the stochastic dynamics might say "The chance of a ball being deflected to the right is 50%", we can get "the typical behavior for a ball hitting a pin is that it gets deflected to the right 50% of the time", and so on. We have gotten our doppelgänger sentences by analysis of a *determinstic* dynamics, in a particular set-up, without needing to choose a *particular* probability distribution over the initial states. We might choose a particular probability measure to do the analysis (or more precisely, to give our plausibility argument), but if the analysis is correct, then any other absolutely continuous measure would work just as well. So we are relieved of the worry of *justifying* a particular choice for the measure that represents the possible initial states. We only have to feel comfortable that it represents a reasonable choice of *sets of measure one and zero*.

All of this seems like magic. We seem to be getting probability out of a deterministic theory without putting anything as detailed as a particular probability measure over initial states in. The probability measure we get out appears to arise almost completely from the deterministic dynamics alone. Isn't this too good to be true?

The first critical observation is that although we have recovered claims about statistical behavior as typical, we have *not* recovered doppelgängers for all of the claims entailed by a stochastic dynamics. Recall that we have only been considering a single ball as it goes down a Galton board. A simple stochastic dynamics would assign a specific probability to *each particular interaction with a pin*. There would be, for example, a 50% chance for the ball to pass pin number 8 to the right. It is from the agglomeration of all of these particular probabilities that the simple probabilistic dynamics would entail a high probability, tending to one, for the long-term proportion of deflections to either side to be 50%. But there is *no typicality statement at all* about the behavior of the ball at the pin 8. It does not typically go to the right, nor typically go to left, and, of course, a single ball never goes 50% to the right *at any particular pin*. So there are many specifiable behaviors to which a stochastic dynamics will assign probabilities, but for which the typicality analysis will have nothing to say. This is not really surprising: intuitively, the typical behaviors (given a deterministic dynamics) will correspond to behaviors assigned very high probabilities by a stochastic theory. And *at pin number 8* the ball is not highly likely to do anything in particular.

Has the typicality analysis lost something important by not having probabilistic claims about particular events? The basic *ontology* obviously does differ from stochastic theory, which contains such claims, but does this loss make any difference to scientific practice?

There is no evident sense in which it does. For although a stochastic dynamics can *assign* a probability to a particular event, the stochastic theory can only be *tested*, and therefore accepted or rejected, by looking at large classes of events that display empirical statistics. Suppose that the stochastic dynamics does assign a probability of 50% for the ball going right at pin 8. Still, when the ball reaches pin 8, it will simply either go right or go left. We only *test* the stochastic theory by seeing the long-term statistics for what the ball does: how often, in a long run, it goes right or left, and whether the deflections to the right and left are randomly distributed according to some statistical test. *But all of this testable behavior, on the basis of which we accept or reject a stochastic theory, could be shown to be typical behavior in a deterministic theory*. This is, indeed, what we expect for the Galton board: not only will 50% deflection to the right be typical, the typical run down the board will pass every statistical test that can be devised for “random” behavior. The typical behavior in the deterministic case will be indistinguishable from behavior assigned probability one by the stochastic dynamics.

If we want to associate *empirical statistics* with pin 8, we obviously need to run *many* balls down the board, not just one. Now we can look and see if about 50% of them get deflected to the right at that particular pin. But the generalization to this case is straightforward: the initial state of the system would now represent the initial positions of *all* the balls. If we have N balls, then the starting points of all the balls would be represented by a point in an N-dimensional configuration space (each ball has one degree of freedom in this analysis). And what we would like to show now is that typical behavior relative to *this* space is that about half the balls are deflected to the right at pin 8 and half to the left. In the infinite case, we hope to show that for a set of measure 1, the frequency of right-deflections at pin 8 is 50%.

What happens when we focus on a finite board, a board of, e.g., 20,000 rows of pins? We can approximate the same logical structure by the use of epsilonics. In such a case, no frequency will be typical in sense defined above. But it will be the case that, relative to the flat measure, *most* of the initial states generate frequencies *near* .5. Let’s choose a small ε, say .00000001. And let’s say that a behavior is typical if it is displayed by a set of initial states that have measure $1 - \varepsilon$ with respect to the flat measure. Then for some δ, it will be typical for the frequency to be $.5 \pm \delta$. The smaller we set ε, the larger we have to set δ, but if the number of rows is long enough, both ε and δ can be made quite small. And having found such an ε and δ, again the fine details of the initial probability measure wash out: exactly the same behavior will could as typical relative to an infinitude of probability measures which are not nearly flat. If the choice of the initial measure was just instrumental in defining typical behavior, then any of these other would have worked just as well, and there need be no fact about which is "objectively correct".

Note that for a board that is too short, the analysis may yield nothing at all. If there are only 20 rows, the only thing that can be said to be typical relative to an ε of .00000001 is that the frequency is between 0 and 1, which we already knew. In this sense, the typicality analysis is useful only for the statistical properties of large classes of events. This contrasts with the use of both stochastic dynamics and probability measures over initial states, which yield precise probabilities for every describable individual event.

The same sort of analysis can be made for more sophisticated properties than just the frequency of passing to the right and to the left. We expect the flips of a fair coin to be independent. This implies a host of statistical properties: if we look at sequences of two flips, we expect to get all four possible outcomes about a quarter of the time, and this should hold whether the two flips immediately follow each other or are separated by intervening flips. Similarly, we expect balls on the Galton board to show independence in how they behave from row to row. The analysis would be similar in principle, but much more complicated.

So we go from the infinite to the finite case by adding a tolerance ε to the frequency (which allows us to define the "good" set) and a tolerance δ to the measure of the "good" set, so we no longer expect it to have measure one. These adjustments come at a certain analytical price. In the infinite case, as we saw, we got the sharp result that if the good set had the property we wanted (measure one) with respect to the chosen measure, then it had the very same property with respect to any absolutely continuous measure, and in this sense, the property we want is defined not with respect to a particular measure but with respect to an easily specifiable huge class of measures. But once we make the adjustments for the finite case, things are not quite so neat. For if the "bad" set is now to have a very, very, small but non-zero measure (with respect to the measure we choose to do the analysis), then it will evidently not have a small measure with respect to every measure absolutely continuous with the first: some absolutely continuous measures will make the "bad" set big and the "good" set small. So there will be some greater degree of sensitivity of the result on the particular measure chosen. Still, it might be quite plausible that this sensitivity is small, perhaps even that a measure according to which the "bad" set has large measure is not even easily describable. So while in the infinite case we can say "If you don't like a spatially flat measure, change to another, but don't go nuts: don't assign a finite measure to a set whose flat measure is zero", we now must say "If you don't like a flat measure, change to another, but don't go nuts: don't assign a large measure to a set of very, very tiny flat measure". If you don't like the spatially flat measure, you are free to change it in almost any reasonable way you can think of: the tiny sets will still be tiny and the large sets still large.

Since the structure of these "typicality" explanations has been largely missed in the philosophical literature, let me recap.

One model of objective probabilities that we have is that of an irreducibly stochastic dynamics. Such a dynamics will assign chances to particular transitions: for a single tritium atom, there will be a chance that it will decay over a stated period of time. That chance obtains even though the initial state of the atom-—indeed of the whole universe-—may be given. When one turns to a deterministic dynamics, it is evident that this precise situation cannot obtain: given the complete specification of the universe at a moment, it is determined what will happen at all later times (and usually, at all earlier times), and there seems to be no room for probabilities or chances. That leads to a certain *completely incorrect* train of thought.

The train of thought runs: if we want something that looks like an objective probability here, then first of all we must not be dealing with predictions from *complete* descriptions of the universe at a moment, we must instead be dealing with predictions from *merely partial* descriptions (e.g. macrodescriptions, or, in Bohm's theory, descriptions at the level of the universal wave-function, or of an effective wave-function). These partial descriptions must leave something out: the exact microstate, or the particle positions. And the fundamental dynamics is defined only for objects at the fine level of description. So (this is where the train goes off the tracks), the only way one gets probabilistic predictions is to supplement the partial description with a *probability measure* over its completions, a probability measure over the microstates compatible with the macrodescription, or over the particle positions compatible with the wavefunction. And once one has such a probability measure (Lebesgue measure, say, or Ψ-squared), then it is trivial to get everything one gets from a stochastic dynamics: given any partial description of a situation, such as “this coin is tossed” or “two boxes of gas at such-and-such temperatures are brought into thermal contact”, one gets a probabilistic prediction for what the result will be: just take the probability measure over the compatible completions of the partial description and let the dynamics evolve that into a probability measure over the possible outcomes. Notice that *if this is what were going on, the method would automatically assign probabilities to particular individual events, given a partial description of them*. In this way, one would get something that looked like a stochastic dynamics which also assigns probabilities to particular individual trials given a (complete) description of the initial set-up.

We have not done anything like this at all. We have not even attempted to assign probabilities to particular individual events—a particular flip of a particular coin, say, or the deflection of the ball on the Galton board at a particular row. There is no “typical” behavior, for example, for what happens at the eighth pin on the board: “typical” behavior, by definition, is behavior displayed by *most* of the possible initial states, and there is nothing to say about how most of those states behave at that pin. What *is* typical is that the long-term frequencies of deflections in each direction are about 50%: this is true for most initial states.

Of course, a stochastic dynamics, as I have been conceiving it assigns a very high probability for long-term frequencies near 50%. It does so by assigning probabilities for particular events (deflection at the first, at the second pin, at third pin, etc.), treating those events as independent, then entailing probabilities for long sequences of particular events, and hence probabilities for long-term frequencies. The approach to deterministic chances we have been considering simply does not follow that order of analysis: rather, the long-term frequencies are shown to be typical without doing anything like assigning a probability to an individual trial. They are typical because most of the initial conditions entail them.

It is true that the "most" in the last sentence requires *something like* a measure over the space of possible initial states, but that measure *is not* being used to "fill in" the missing parts of a partial description. What we get at the end is not like a stochastic dynamics in one sense: it will not assign anything like a "probability" to *particular* events. But on reflection, this is seen not to be a methodological deficit. For the empirical significance of probabilistic theories—the empirical facts that provide evidence for them and which are explained by them—are never single particular facts. They are rather collections of particular facts, collections of "trials", all taken to be similar in some way, which display empirical statistics: frequencies of outcomes, with the different particular outcomes randomly distributed among the trials. It is *these* that first call for an explanation, *these* that suggest to us, in the first place, a probabilistic treatment of the phenomena, and *these* that allow for the testing of theories. And while a stochastic dynamics deals with these empirical distributions by *implying they are highly likely* (in a sense of "likely" such that there is a fact about how likely each individual event was), a deterministic dynamics rather can be shown to imply that these same distributions are *typical*, without treating the individual cases first and deriving probabilities for the collections from them. Indeed, the order of analysis is just the other way around: if one wants to say that the chance of a *particular* coin flip coming heads is 50%, the typicality analysis can only *make sense* of this as a shorthand for "this coin is a member of a naturally defined collective of coins for which a frequency of around 50% is typical, and there is no reason to think this coin is special". Absent such a collective, the typicality analysis would not get off the ground.[8]

Put another way, a stochastic theory can assign a probability of 50% to a single event without there ever being anything in the universe but that one event. (Of course, if only that event ever occurred, no one could have *good grounds* for thinking it had that probability, but that is an epistemic matter, not an ontic one.) No typicality analysis could

[8] Note here that we are concerned with probabilistic claims that involve probabilities not near zero or one. Even if there were only one box of gas in the universe, if the gas started in one corner, one could say that typical behavior is for it to expand to fill the box: most of the initial states lead to that result.

yield an analogous result. But the fundamental differences between the approaches, evident at the individual level, make essentially no difference at all to how one *uses* the theories to account for empirical statistics, because those situations must contain collectives with empirical frequencies. A stochastic dynamics can imply that the observed frequencies, the random distribution of particular results, etc., are all *highly probable*. The typicality analysis can show that these very same frequencies, distributions of results, etc., are *typical*. And the basic suggestion is that each of these sorts of treatments of the empirical distribution is equally acceptable as a scientific account of the phenomena. In exactly this sense, it appears that a deterministic theory can give you everything important that a stochastic one can: indeed, no *observable behavior* could indicate that one treatment or the other was more appropriate.

The "add a probability measure to fill out the partial description" strategy, if it could be made to work, would provide a much broader set of probabilistic predictions than the "typicality" strategy. It would, in principle, provide probabilistic predictions for any clearly articulated question. There would be a presently defined *physical* probability, for example, that the Cubs will win the worlds series in 2050 (assuming we had a precisely defined present macrostate of the world). We might not ever know what the probability is, but it would nonetheless exist. But there is no reason at all to believe that the typicality analysis could even be brought to bear on this question.

What the typicality analysis loses in terms of propositions about single events is compensated for by not requiring a commitment to any particular probability measure. Applying the notion of typicality does not require anything nearly as mathematically detailed as a probability measure over initial states. What it requires instead is a rough-and-ready division of sets of initial states into "large" and "small". And not all sets of initial states need to be so categorized: one is only concerned with sets of states that all give rise to the sort of empirical statistical behavior one might observe and record. The main point is that with respect to those sets, many wildly different probability measures will yield exactly the same categorization into "large" and "small". Any particular one of these probability measures would then be not more "objectively correct" than any other, and the *use* of one rather than another to do calculations can be justified on purely pragmatic grounds.

A concrete example of this can serve to illustrate the importance of distinguishing a typicality analysis from some other type of probabilistic analysis. One of the most important examples of a deterministic theory having to account for statistical phenomena is Bohmian mechanics. Since the theory is deterministic, but is also designed to account for the probabilistic predictions of the quantum formalism, this particular problem is very acute. And it is often said that Bohmian mechanics is only able to deliver the probabilistic predictions of quantum theory because it postulates a very particular measure over the possible initial configurations of the universe: the Ψ-squared measure, where Ψ is the initial universal wavefunction. Since the squaring of the wavefunction (N.B.: the wavefuction of a *subsystem* of the universe, not the whole universe) is used in the standard quantum formalism to calculate probabilities, it can seem plausible to say that *Bohmian mechanics can only get the right probabilities out if it puts the right probability measure, viz. the Ψ-squared measure, in at the beginning*. This makes the whole theory look like an *ad hoc* conjuring trick.

The typicality analysis, however, shows that this is *not* what is going on. It is true that the Ψ-squared measure is often used for derivations, just as the spatially flat measure would be used to analyze the Galton board. That is because the Ψ-squared measure plays nicely with the dynamics of the theory: it is equivariant. Use of any other measure for tracking what happens to "big" sets of initial conditions would be mathematically intractable. But at the end of the day, all one sets out to prove is that the typical observed frequencies in large collections of experiments are exactly the frequencies which standard quantum mechanics predicts by ascribing probabilities to individual events. If this analysis goes through for the Ψ-squared measure, *it would just as well go through* (in the limit of infinite repetitions) for any other measure absolutely continuous with Ψ-squared. In particular, there is every reason to believe that if the quantum-mechanical frequencies are typical with respect to the Ψ-squared measure's categorization of "big" and "small" sets of universal initial conditions, they are also typical with respect to a spatially flat, uncorrelated measure's categorization of "big" and "small". In this case, the analysis is not *physically* or *metaphysically* tied to Ψ-squared measure: use of that measure is just a matter of convenience. The only *objective* physical structure that needs to be postulated is the somewhat vague categorization of some sets of initial states into big and small.

Summation

We have surveyed three quite different ways in which probabilities—or at least probability measures associated with talk about *chances* and connected to expectations about *frequencies*—can be defined on a purely physicalistic foundation, with no reference to subjective states. In a stochastic dynamics, the measure is itself part of the foundational physics, and would be postulated to be perfectly well-defined. In the Humean case, the measure might not be uniquely defined, since many such objects, mathematically different, might do equally well at encapsulating information about the Humean mosaic. In the case of deterministic chances, one gets a unique set of empirical frequencies—exactly one set of frequencies in a model will be typical—but this is an analytical consequence of the deterministic dynamics together with something much weaker than a probability measure over initial states. The three approaches all have different ontologies and consequences. The stochastic dynamics postulates something fundamentally "chancy" and entails that the observable world might be highly misleading: if there is bad luck, the chance results might provide better evidence for a false theory than the true one. The Humean approach adds nothing to the ontology, and could make no sense of a "bad luck" scenario: the probabilistic assertions are just ways of compactly conveying information about the actual distribution of particulars in the world. The Humean approach postulates no intrinsic connection between the probability measure and dynamics. Finally, the typicality approach articulates how do make an analytical connection between a deterministic dynamics and a characterization of certain empirical distributions as "typical". There is no guarantee that the typical frequencies will be actual: one might have the dynamics right but again be subject to bad luck, since atypical behavior is not ruled out. But if the actual behavior of the world corresponds to typical behavior (relative to the dynamics), then this appears to provide a physical account of the behavior, and to provide as well further evidence that the postulated dynamics is correct. Typical frequencies can then be used to provide a probability measure over results in individual cases.

None of these methods for introducing a probability measure could be thought of as generating the “measure over propositions” needed for psychology: none, for example, could make any sense of a “probability that string theory is true”. The psychologist, and the normative epistemologist, need such things, or at least need a way to represent different strengths of credibility attached to string theory. But the physicist, *qua* physicist, does not. The physicist is concerned with physical behavior, not with psychology. He can have the tools of objective chance for his job, even if it is not everything that the psychologist or normative epistemologist needs. What the psychologist and normative epistemologist cannot do is insist that *their* use of a probability measure over propositions is the *only* proper meaning of “probability” or “chance”, especially since in this case, the use of a probability measure is not even particularly well-suited for the job.

Bibliography

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

Dürr, D., S. Goldstein and N. Zanghì (1992). Quantum Equilibrium and the Origin of Absolute Uncertainty. *Journal of Statistical Physics* **67**, 843-907.

Earman, John (1986). *A Primer of Determinism*. Dordrecht: Reidel.

Feynman, Richard (1967). *The Character of Physical Law*. Canbridge, Mass.: MIT Press.

Lewis, David (1986). *Philosophical Papers, Volume II*. Oxford: Oxford University Press.

Loewer, Barry (2001). Determinism and Chance. *Studies in History and Philosophy of Modern Physics* 32: 609-620

----------- (2004). David Lewis's Humean Theory of Objective Chance. *Philosophy of Science* 71, No 5: 1115-1125.

Lucas L. and M. Unterweger (2000). Comprehensive Review and Critical Evaluation of the Half-life of Tritium. *Journal of Research of the National Institute of Standards and Technology*, volume 105, Number 4, pp. 541-149

Maudlin, Tim (2007). What Could Be Objective About Probabilities? *Studies in the History and Philosophy of Modern Physics* **38,** 275-91