From: Chapter 9; Approaches to Probabilistic Explanation

9.2 Varieties of Probabilistic Explanation: Examples

Quantum Mechanics: Half-Life of an Element The half-life of radium-226 is 1620 years, meaning that about one half of any given sample of this isotope of radium will have decayed by the end of a 1620 year period. Why? Some sophisticated quantum mechanics determines both a particular probability that any given atom of radium-226 decays in a given time period, and assures us that the decays are stochastically independent. The probability that a given atom decays in a 1620 year time period is almost exactly one-half; given independence, then, and the law of large numbers, there is a very high probability—in a sample visible to the eye, only negligibly less than one—that about one half of the atoms in a given sample will have decayed when 1620 years have elapsed.[2]

Quantum Mechanics: Tunneling Radium-226 decays by emitting (along with some electromagnetic radiation) an alpha particle, a bundle of two neutrons and two protons. Classical models of the atom cannot explain alpha decay, as the emitted alpha particle must jump a potential barrier—a wall of force, if you like—so high that it would have negative kinetic energy at the apex of its leap. But radium atoms emit alpha particles all the same.

Quantum mechanics explains alpha decay by showing that there is a probability—a very small probability, but a probability nevertheless—that an alpha

2. This is the standard way of talking probabilistically about radioactive decay. It is not entirely in accord with the standard way of talking about probability in quantum mechanics, the "collapse interpretation", on which atoms cannot be said to have decayed until the decay is measured. It can still be said that, at the end of 1620 years, there is a very high probability that a measurement of a radium sample will show that almost exactly half of the sample has decayed, which is sufficient for my purposes here.

particle will jump the wall, or rather, in the preferred metaphor of quantum physicists, that it will tunnel through the wall.[3] (These are, of course, the same probabilities that, taken en masse using the law of large numbers, explain the half-life of radium. They are small, then, only when the period of time that the radium atom is observed is small, say, one hour.) It is the existence of this small probability that is taken to explain a given atom's alpha decay, when it occurs, or perhaps better, to explain how alpha decay is possible.

Sociology: Crime and the Family African American teenagers are more likely than their peers to commit violent crimes. Many sociologists cite, as a partial explanation of this fact, the relatively higher rate of young, single parents in African American homes. Having a young, single parent is said to increase the probability of a teenager's engaging in violent crime; single parenthood of this sort therefore explains, in part, the higher rate of violent crime among African American teenagers.

Medical Science: John Jones Recovers Swiftly John Jones has a streptococcus infection. He is given penicillin. He recovers within a week. The probability of such a swift recovery without penicillin is 10%; the probability with penicillin is 90% (Hempel 1965a, §3). It seems that the administration of the penicillin, then, explains the speed of Jones's recovery.

Statistical Mechanics: Nature's Abhorrence of a Vacuum When a vacuum is created—say, all air is evacuated from a bell jar—and then opened to the surrounding air, as when a valve is opened in the side of the bell jar, the air rushes to fill the vacuum. Statistical mechanics, in conjunction with the kinetic theory of gases, explains this behavior as follows.

The vacuum is empty space. The surrounding area is filled with gas particles. These gas particles move around the space available to them at random. (A more precise probabilistic characterization of their movements is

3. For a cautionary comment about this way of talking, see note 2.

possible, but is unnecessary here.) A mathematical argument shows that randomly moving particles will tend, over time, to distribute themselves evenly in the space available, if that space is bounded. When the bell jar's valve is opened, the empty space inside the jar becomes available: the gas particles are no longer prevented from entering the jar. Thus, over time, the gas particles will distribute themselves inside the jar with a density equal to the density of the particles outside the jar. To the human observer, this redistribution takes the form of a rush of air into the jar.

To better understand the role of probability in this explanation, let me spell out the three main steps of the mathematical argument establishing a tendency to even distribution.

1. The probabilistic premise of the argument is that gas particles are engaged in a *random walk* through the space available to them.

2. A mathematical theorem shows that a particle engaged in a random walk through a bounded space will, over time, be equally likely to be found anywhere in the space. That is, after a certain amount of time has elapsed (determined by the pace of the walk and the shape and size of the space), the probability distribution over the particle's position will be uniform, meaning that the particle is equally likely to be found anywhere in the space available.

3. Because there are so many particles in a gas, and each is equally likely to be found anywhere in the space available, the law of large numbers can be applied to deduce a very high probability that the gas will be distributed evenly through space.

The tendency for gas to rush into an evacuated space, then, is a probabilistic tendency: there is a very high probability that the space will be filled, but a slender chance that it will remain empty.[4]

---

4. I have given an explanation of the gas's behavior in the Boltzmannian vein; an expla-

Evolutionary Genetics: Homozygosity in Elephant Seals The explanation of elephant seal homozygosity was presented in section 8.13 as an example of an explanation that places great importance on a mathematical fact. I will use it here as a paradigm of an explanation that places great importance on a probabilistic fact. That fact, recall, the cornerstone of the theory of drift, is that extinction through drift is more likely the smaller the relevant population. Because elephant seals were hunted to near extinction, their population was very small for long enough that most alleles that survived the hunting drifted to extinction. For this reason, every known loci in the elephant seal genome is homozygous. Homozygosity in elephant seals is explained, then, by the relatively high probability of allele extinction during the period in which the population of elephant seals was very small.

Evolutionary Ecology: Finch Sex and Famine On the Galapagos islands, the ratio of male to female finches of the species *Geospiza fortis* after a drought year is much higher than the 50:50 ratio found in normal years (Grant 1986). This is because

1. As a drought goes on, a higher and higher proportion of the edible seeds remaining on the islands are large and difficult to crack,
2. Larger-bodied, deeper-beaked birds have an easier time with large, hard seeds,
3. The probability of a finch's surviving a drought increases with the size of its body and the depth of its beak.
4. Males are on average 5% larger than females, thus males are on average more likely to survive a drought than females.

---

nation in the Gibbsian vein directs attention away from probability distributions over the positions of individual particles and towards probability distributions over the configuration of the gas as a whole. The two explanations have in common the attribution of a very high probability to the event of the gas's filling the vacuum. For a philosophical perspective on these and many other issues in the foundations of statistical mechanics, see Sklar (1993).

Premise (3) is probabilistic because some small finches survive a drought and some large finches do not.

This is an example of the sort of ecological explanation that provides the foundation for much Darwinian explanation in evolutionary biology. Ecological concerns show that, in a given environment, organisms with a certain trait—in this case, large body size and beak depth—have greater fitness, that is, roughly, a higher probability of survival or reproduction. The law of large numbers is then invoked to conclude that the representation of that trait in the population will increase as long as the environment persists. Usually, the environment changes slowly if at all, and the advantageous trait is fixed: it completely replaces all other competing traits in the population.

9.3 Varieties of Probabilistic Explanation: Commentary

The Explanandum In some explanations, the explanandum is a single event, while in others, it is a regularity. It is a single event or state of affairs in the tunneling, swift recovery, and elephant seal homozygosity explanations. It is a regularity in the half-life, crime, vacuum, and finch sex ratio explanations.

There is, however, a close relationship between the probabilistic explanation of single events and of regularities. The same probabilities that are used to explain single events are also used, in conjunction with the law or large numbers, to explain regularities that consist in repeated occurrences of events of the same sort. For example, the probability used to explain a particular occurrence of alpha decay also explains the overall rate of decay, as in the half-life explanation. Equally, the probabilistic reasoning used to explain why gases in general rush to fill vacuums is also used to explain why some particular gas rushes to fill some particular vacuum. (It is this observation that inspires the kairetic account of theoretical explanation, both deterministic and probabilistic.)

The Uses of Probabilistic Explanation There are at least three different reasons for using probability in these explanations.

First, in the half-life and tunneling explanations, the quantum probabilities involved are, as far as we know, irreducible, so that there is no deterministic explanation: the probabilistic explanations are the only explanations we have.

Second, in the crime and swift recovery explanations, the use of probability is due in part to our ignorance of all of the relevant processes and information. Suppose, for example, that there are two equally common strains of streptococcus, one of which responds to penicillin 80% of the time, the other 100% of the time. If you knew that John Jones's infection were of the latter variety, you would give it a deterministic explanation. It is only because of your ignorance about the strain that you turn to probability.

A similar but more subtle point can be made by supposing that the more susceptible strain is eradicated swiftly by penicillin 99% of the time. More knowledge, in this case, does not lead to a deterministic explanation, but to a *different* probabilistic explanation, different because it cites a different probability. Here, perhaps, probabilistic explanation is unavoidable, but we give the particular probabilistic explanation we do out of ignorance.

The same observation can be made about explanations of regularities, as the crime case shows. Some kinds of single parent may not increase the tendency to violent crime at all, some may increase it only a little, and for some the increase, if any, may depend on other environmental factors. It is ignorance alone, it seems, that leads scientists to cite a single probability, and not a complex statistical or deterministic relationship, in the explanation of the crime rate.

The third and final reason for using probabilistic explanation is illustrated by the examples from statistical mechanics and evolutionary biology. In complex systems such as gases and ecosystems, probabilistic reasoning allows you to ignore the causal intricacies of the interactions between the systems' many parts, while still deriving a near-one probability for the phenomenon of interest (Strevens 2003b, 2004b). Probability allows you to see the big picture without being distracted by unnecessary details. It is this explanatory use of probability that interests me most.

The Varieties of Explanatory Probability Roughly paralleling the three uses of probabilistic explanation sketched in the last section are three different kinds of probability appearing in my example explanations, which I dub simple probability, quasi-probability, and complex probability.

A *simple probability* is a metaphysically irreducible probability. The leading candidates in the world as we know it are the probabilities of quantum mechanics. Supposing that quantum mechanical probabilities are indeed irreducible—that there is no deterministic process underlying the processes that quantum mechanics declares probabilistic—then the probabilities appearing in the half-life and tunneling explanations above are simple probabilities.

A *complex probability* is a metaphysically reducible probability. Complex probabilities are found wherever a true scientific theory describes a process in exact probabilistic terms that is deep down deterministic. (If the probabilistic terms are inexact, the theory may be quasi-probabilistic.) A simple example is the one-half probability that a tossed coin lands heads. More sophisticated, and controversial, examples are the probabilities of statistical mechanics and of evolutionary ecology.

The controversy comes from two quarters: first, the metaphysical view that only irreducible probabilities are "real", and second, the possibility that the probabilities of statistical mechanics are higher level manifestations of irreducible quantum mechanical probabilities.

The first consideration—the reality of complex probabilities—is, for my purposes, beside the point: what is important is whether complex probabilities are explanatory, and science clearly treats them as such. (Still, as we will see in section 9.6, there are those who deny the explanatory power of complex probabilities precisely on the grounds that they are not "real".)

What if the probabilities of statistical mechanics and evolutionary biology are just simple probabilities differently expressed? Observe that these probabilities were treated as explanatory even in the great age of determinism, the nineteenth century,[5] and are treated as explanatory even today when we are unsure of their quantum provenance. This is enough, I think, to show that their explanatory power is independent of their metaphysical roots. I propose to treat them as definitely non-simple, so as to have some clean examples of complex probability for the sake of the argument.

I expect neither of these maneuvers to mollify the implacable foes of complex probability. My more substantive contribution to this debate will be a demonstration, in chapters ten and eleven, that the kairetic account of explanation *demands* that complex probabilities be posited and used to explain certain behaviors of certain deterministic systems. Complex probabilities not only *can* be used to explain; in some circumstances they *must* be used to explain.

My third kind of probability is *quasi-probability*. Quasi-probabilities are explanatorily potent, and are sometimes featured in the very best explanation of an explanandum, but they do not have exact values. They may be thought of as imperfectly formed complex probabilities, in a sense that will become much clearer when they are discussed at length in section 10.3. A quasi-probability is a kind of statistical amalgam, like the probabilities cited in the violent crime and swift recovery explanations above; however, not every such amalgam is a quasi-probability. I will argue in chapter ten that it is explanatory concerns that distinguish the quasi-probabilities from less exalted amalgams: it is in those special circumstances where an amalgam provides the best explanation of the explanandum that the amalgam can be counted as a genuine quasi-probability, a permanent part of the scientific picture of the world rather than a tool to be discarded as soon as more information arrives.

5. Here I read Darwinian arguments as involving a kind of informal probabilistic reasoning.

There may be probabilities that are not merely quasi-probabilities, but that are neither simple nor complex. They will depend in complicated ways on both metaphysically irreducible probabilities and on the kinds of deterministic processes that underlie complex probability. I baptized these *simplex probabilities* in Strevens (2003b); they are discussed further in that book. The treatment of simplex probability is involved, but brings no new insights about the nature of probabilistic explanation, so I will not mention it again.

I said above that there is a rough parallel between the three applications of probabilistic explanation, described in the previous subsection, and the three kinds of explanatory probability. Let me explain. When we turn to probabilistic explanation because no deterministic explanation is available, we cite simple (or simplex) probabilities. When we resort to probabilistic explanation out of ignorance, we use probabilities that are amalgams of simple or complex probabilities. If we are lucky, these probabilities are quasi-probabilities, and our explanations will be enduring. When we use probabilistic explanation to bypass complex interactions that ultimately make little difference to the explanandum, we typically use complex probabilities, or so I argue in Strevens (2003b). Possibly, however, we use simplex probabilities or even quasi-probabilities—the parallel is not perfect.

The Forms of Probabilistic Explanation There will be much more to say on this topic later (sections 9.5 and 9.6). Let me just note that, with respect to the elitism/egalitarianism debate, there seem to be almost as many probabilistic relations between an explainer and an explanandum in this small sample of scientific explanations as have been suggested in the entire explanation literature.

In some cases, an event or regularity is explained by citing a high probability for the same: the half-life and vacuum cases are examples. In some cases, citing a very low probability does the explaining: tunneling provides a paradigm. In still other cases, the explanation proceeds by citing a factor that greatly increases the probability of the explanandum, as in the quick recovery, or that merely increases the probability to some degree, as in the case of violent crime.

The one probabilistic relation sometimes claimed to be explanatory that is not represented in the examples I have chosen is the relation of *decreasing* the probability of the explanandum. The explanatory status of the probability-decreasing relation will be discussed further in section 9.5.

9.4 Approaches to Probabilistic Explanation

Having tabled some evidence, let me return to the top-down approach to explanation, next sketching very briefly some theories of probabilistic event explanation influential over the last forty years, beginning with Hempel.

Nomic Expectability On Hempel's nomic expectability account of explanation, an explanation is a law-involving argument that provides good reason to believe that the explanandum occurred. In the D-N account the argument is deductive; in Hempel's I-S account of probabilistic explanation (Hempel 1965a, §3), it is inductive. According to Hempel, then, a probabilistic explanation is a sound inductive argument with the conclusion that the explanandum occurred. The requirement of soundness imposes two distinctive conditions on probabilistic explanations, according to Hempel:

1. Probabilistic explanations must satisfy a near-equivalent of the requirement of total evidence. That is, they must mention every factor known to affect the probability of the explanandum.[6] It follows that the good-

6. In the I-S account of explanation, the requirement of total evidence is amended to become what Hempel calls the *requirement of maximal specificity*. The reason for the amendment is that the total evidence principle requires that you take into account all relevant information known to you at the time the argument is formulated, which would result in your having to include as a premise of the argument the proposition that the explanandum actually occurred. This would make the argument deductive, and rather unexplanatory. It is not as easy as it might appear to formulate an appropriate version of the maximal specificity requirement.

ness of a given I-S explanation is relativized to the knowledge situation of the explanation's beneficiary.

2. The premises of a probabilistic explanation must confer a high probability on the explanandum—whatever probability is sufficient, according to the canons of inductive logic, for acceptance of a conclusion probabilified to that degree, but at any rate, greater than one half.

Objections to Hempel's account have largely focused on these two conditions.

Dispositional Accounts The idea motivating a dispositional account of probabilistic explanation is that probabilities are a kind of disposition, like fragility and solubility, and ought to feature in explanations in just the way that dispositions do (Coffa 1974; Fetzer 1974). As such, probabilities ought to be thought of as causal properties responsible for bringing about the outcomes to which they are attached. Dispositional accounts belong, then, to the causal tradition in explanation. Like other causal approaches, they are not committed to any particular metaphysics of dispositions: a dispositionalist could also be an metaphysical empiricist, given an empiricist account of dispositions.

Probabilistic Relevance A probabilistic relevance account holds that an event E is explained by those factors that change the probability of E, either increasing it or decreasing it (Salmon 1970; Railton 1978; Humphreys 1989). An alternative version holds that only factors that increase the probability are explainers (van Fraassen 1980; Sober 1984).

Probabilistic relevance accounts can be given as part of either a pattern subsumption account of explanation (possibly Salmon's position in Salmon (1970), before his conversion to the causal approach), or as part of a causal account of explanation. In the causal version, the probabilistically relevant factors are held to be causally related to the explanandum in virtue of their contribution to the probability. It is left open whether or not the probability itself is a causal property. Salmon (1990a) holds that it is indeed causal—

it is a kind of disposition. Humphreys (1989, 141) disagrees, stating somewhat enigmatically that "chance is . . . literally nothing", enigmatically because what Humphreys regards as the true explainers—the probabilistically relevant factors—owe their explanatory power entirely to their effect on this "nothing".[7]

Railton's D-N-P Account Railton's D-N-P account of explanation confines itself to a thesis about probabilistic explanation's formal requirements. A probabilistic explanation, according to the D-N-P account, is a complete statement of all the facts and laws relevant to the probability of the explanandum, a (deductive) derivation of the probability from those facts and laws, and a note to the effect that the explanandum occurred. So far, the formal requirements seem compatible with any major view about the explanatory relation.

Railton adds, however, two more elements to his account. First, he specifies that the size of the probability attached to the explanandum is irrelevant to the force of the explanation. This rules out an expectability interpretation of the D-N-P account. Second, as already noted above, he makes a deep distinction between probabilistic explanation and explanation in deterministic systems, referring to the latter alone as causal explanation. This rules out, apparently, a causal interpretation of the D-N-P account, and with it, perhaps, a dispositional interpretation (but see note 1).

Two Auxiliary Debates I will focus the rest of my discussion on two auxiliary debates from the probabilistic explanation literature. Both debates can

7. A charitable interpretation of Humphreys, based on his appeal to the structure of linear causal models, would see him as separating the process of the probabilistic production of an event into two parts: a deterministic part, in which the probabilistically relevant factors directly bring about the outcome, and a part in which the deterministic causal process is blurred slightly, due to Nature's inability to pull off the bringing about entirely according to specifications. It is this latter part of the process, I think, that Humphreys considers the mere nothing, a moment of incontinence on Nature's part. This view of things makes the mistake, I think, of attributing to probability only a part of what probability is responsible for—the standard deviation, as it were, but not the mean.

be joined in the top-down fashion, by taking a thesis about the explanatory relation as the major premise. But I will, at this stage, take the bottom-up approach, commenting on the debates in the light of the examples of probabilistic explanation discussed in section 9.2. A top-down resolution of many aspects of the debates, founded on the kairetic account's explanatory relation, will follow in chapters ten and eleven.

The debates, in the order of discussion, concern the following questions:

1. Does the size of a probability or the size of a change in probability make a difference to the quality of the explanation? (This is what I refer to as the elitism/egalitarianism debate.)

2. Can there be a probabilistic explanation of a deterministically produced event?

10 | Kairetic Explanation of Frequencies

I want to show that the best explanation of a deterministically produced event or set of events is sometimes probabilistic. In this chapter, I consider the explanation of frequencies of events within a particular sample population, such as the frequency of alpha decay in a given sample of radium or the preponderance of large over small Galapagos finches after a particular drought. In the next chapter, I consider the explanation of single events, such as a particular alpha decay or John Jones' swift recovery. Theoretical explanation, including the explanation of ongoing regularities such as the fact that gases almost *always* occupy an available vacuum or the fact that droughts almost *always* result in a preponderance of large over small finches, is discussed in section 11.6.

I have divided the explanation of frequencies and single events into two separate topics not because there is any deep difference between the explanation of single events and the explanation of frequencies within a sample, but rather for expository reasons: some issues are more easily discussed using one kind of explanandum, some using the other. The frequency chapter will be concerned with the fundamental reasons that deterministically produced explananda are best understood probabilistically. The single event chapter will be concerned with the subtle differences between explaining and difference-making, and in particular, with the observation that some probability-raisers and some difference-makers are not explainers.

10.1 Probabilistic Explanation in Deterministic Systems

I have tentatively distinguished two kinds of probabilistic explanation in deterministic systems: explanation using complex probability, and explanation using quasi-probability. These will be discussed separately in sections 10.2 and 10.3, and the notions of complex probability and quasi-probability will be explicated and given a proper foundation along the way. Before I plunge into the details, though, let me explain what complex probabilistic and quasi-probabilistic explanation have in common.

Probabilistic explanation is compulsory in deterministic systems when a probabilistic model of the causal production of the explanandum has greater combined generality and accuracy than any deterministic model. The probabilistic model will never be as accurate as the deterministic model, but in many cases, it is so much more general—so much more abstract, so much less detailed—that the generality gained by going probabilistic easily compensates for the loss of accuracy. If you take a deterministic model as your starting point for the explanation of some explanandum *E*, then, the optimizing procedure will convert it into a probabilistic model for *E*. The deterministic details dropped in the process are thereby declared explanatorily irrelevant.

In order to show that the optimizing procedure has this power to transmute the deterministic into the probabilistic, I must show how to quantify the accuracy and generality of probabilistic models.

Accuracy I have already laid down the key doctrine of accuracy: the accuracy of a probabilistic model for an event *E* is the probability that the model assigns to *E*. A deterministic model, then, has an accuracy of 1. If a model ascribes a 0.9 probability to a phenomenon, the accuracy of the model is 0.9. And so on. When a model is instantiated by systems that ascribe different probabilities to its target, the accuracy of the model is the lowest such probability.

Why accept a probabilistic definition of accuracy? Intuitively, you may think of a model's accuracy, so defined, as the expected frequency with which the systems instantiating the model produce the explanandum—that is, the rate at which the systems can be expected to hit the explanatory target.

But in the kairetic account, accuracy is only a means to an end. Does this particular conception of accuracy, when embedded in the kairetic criterion for relevance, capture our intuitions as to which factors are relevant to probabilistic production, and how the relevant factors figure in probabilistic explanations? This, as you will see in chapter eleven, turns out to be a complex question. In the end the answer is, overwhelmingly, *yes*.

Generality How to compare the generality of a probabilistic model with that of a deterministic model? Consider two models for a given target that differ only in the following respect: the deterministic model states that an event *C* occurred, while the probabilistic model states only that *C* had a certain probability of occurring. (The size of the probability is irrelevant.) Which model is more general?

The setup of neither model entails the setup of the other. That *C* has a certain probability of occurring does not, of course, entail that *C* in fact occurs; but equally, that *C* occurs does not entail a particular value for the probability of *C*'s occurrence. Each model tells you something about *C* that the other does not.

For my purposes in what follows, I need only the following doctrine about generality: the generality of the above two models is approximately equal. Moving from the deterministic to the probabilistic model, then, does not cost much, if anything, in the way of generality.

It will cost something in accuracy, assuming that *C* is essentially involved in entailing the models' common target *E*: the deterministic model entails *E*, whereas the probabilistic model entails only a certain probability for *E*, equal to the probability of *C*. Thus moving from a deterministic to a probabilistic model is, all things being equal, not sanctioned by the optimizing procedure.

If, however, the move to a probability distribution over *C* allows much other causal detail to be discarded, then there is generality to be gained by making the move. If the cost in accuracy is not too high, the probabilistic model will be rated better than the deterministic model by the optimizing procedure. It is in these circumstances that the best explanation of a deterministically produced event is probabilistic, or so I will argue in what follows.

Two remarks. First, the move from a deterministic model to a probabilistic model always incurs a cost in accuracy. In the discussion of accuracy/generality tradeoffs in section 5.2, I concluded that the factors discarded when such a tradeoff is made are difference-makers that are not explanatorily relevant. This doctrine carries over to the probabilistic case: the factors discarded in the move from the best deterministic model to a probabilistic model are strictly speaking difference-makers, but they are explanatorily irrelevant. This will be an important topic in the coming discussion; see in particular section 11.3.

Second, before I proceed there is one technical nicety to be finessed. I claim that the explanatory kernel corresponding to a deterministic model $\mathcal{M}$ is sometimes a probabilistic model $\mathcal{P}$. The criteria for kernel construction laid down in chapter three require that, in this case, $\mathcal{P}$ generate $\mathcal{M}$. But since $\mathcal{M}$ does not entail $\mathcal{P}$, the generation relation does not obtain.

The solution is to expand the generation relation: say that a set of statements $\mathcal{S}$ *probabilistically generates* another set of statements $\mathcal{S}'$ if $\mathcal{S}'$ asserts, and only asserts, the obtaining of one or more states of affairs, and $\mathcal{S}$ asserts, and only asserts, probability distributions over these states of affairs. Then $\mathcal{P}$ *probabilistically generates* $\mathcal{M}$ if the setups of the two models can be divided into two parts, such that the first part of $\mathcal{P}$'s setup generates the first part of $\mathcal{M}$'s setup and the second part of $\mathcal{P}$'s setup probabilistically generates the second part of $\mathcal{M}$'s setup.

The optimizing procedure is then revised so that, for $\mathcal{P}$ to be a kernel for $\mathcal{M}$, either $\mathcal{P}$ must generate $\mathcal{M}$ or $\mathcal{P}$ must probabilistically generate $\mathcal{M}$. (The second clause, of course, subsumes the first.)

Note that I could, but that I have not, redefined generality using this new notion. That is, I have not stipulated that $\mathcal{P}$ is *more abstract* than $\mathcal{M}$ if it probabilistically generates $\mathcal{M}$, so that a move from $\mathcal{M}$ to $\mathcal{P}$ brings an increase in generality. It might be interesting to explore the consequences of this amendment to the kairetic account.

10.2 Explanation with Complex Probability

10.21 Complex Probability

A complex probability is a metaphysically reducible probability. But under what circumstances does it make sense for science to posit complex probabilities? More simply, where do complex probabilities exist? Without some preliminary answer to this question, a discussion of the explanatory uses of complex probability may seem in vain.

One solution to the problem is characterization by example: complex probabilities include the probabilities attached to simple gambling devices such as tossed dice, the probabilities of statistical mechanics, and the closely related probabilities of evolutionary ecology and population genetics.[1]

It will be almost as easy, and will rather better illuminate one of the central contentions of this book, to lay down a definition of complex probability.

Some groundwork: a set of events, such as the outcomes of a series of die tosses, is *probabilistically patterned* just in case the events have the aspect typical of a probabilistic process. There are many kinds of probabilistic processes, for which reason this definition is rather open-ended, but let me fix on one kind of process in particular, a Bernoulli process.

Die tosses and roulette games are examples of Bernoulli processes. The pattern typical of a Bernoulli process's outcomes combines two elements, one of short term disorder and one of long term order. The long term order resides in the tendency of outcomes of a Bernoulli process to occur, in the long run,

1. I put aside the possibility that quantum probabilities "percolate up", making my exemplary probabilities nothing more than higher level manifestations of quantum probabilities.

with a given frequency. One half of a long series of coin tosses will tend to be heads; one half of a long series of die tosses will tend to be sixes, and so on. The short term disorder lies in a complete lack of correlation between one event and the next in the series: once the long run frequency of sixes among die tosses is known, knowing the result of some particular die toss will tell you nothing about the results of any of the others.

A *Bernoulli probabilistic pattern*, then, is a pattern that has these aspects of long term order and short term disorder. If you like, it is a pattern that will pass a statistical test for the existence of an underlying Bernoulli process. Other probabilistic processes, such as Poisson processes and random walks, have their own characteristic patterns.

When science finds a probabilistic pattern in nature, it usually posits a probabilistic process of the corresponding sort, with appropriate values for the probabilities. (In the case of a Bernoulli process, the probabilities are set more or less equal to the frequencies.) If the system producing the pattern is at root deterministic, then it is a complex probability distribution that is thereby posited. When is science right in supposing the existence of complex probabilities? Just when, I propose, the correct explanation of the probabilistic pattern turns out to be a probabilistic model, that is, just when the model for the pattern that offers the best combination of accuracy and generality is probabilistic.

To repeat my claim in a more orderly way, a complex probability distribution over a set of possible outcomes exists in a system where:

1. The outcomes are produced deterministically,
2. The outcomes are probabilistically patterned, and
3. The probabilistic patterns are best explained by a probabilistic model.

It is the theory of explanation, then, that determines what complex probabilities exist. The definition of complex probability is not complete without an account of probabilistic explanation in deterministic systems that does not assume the existence of complex probability. That is exactly what I will now provide.

Three remarks. First, the above definition does not clearly distinguish complex probability and quasi-probability. The difference between the two will be discussed in section 10.3.

Second, there are at least some deterministically produced probabilistic patterns that are best explained non-probabilistically, such as a pattern of ones and zeroes carefully constructed by hand to pass the appropriate statistical tests. The explanatory test, then, has some bite: a probabilistic pattern does not always signify a complex probability distribution.

Three, an aim of this approach to complex probability is to use the criteria for explanatory goodness to solve Fetzer's objectivity problem (section 9.6). Fetzer's worry, recall, was that there is no set of criteria for deciding on a single complex probability to be used in explaining a given deterministically produced event. My aim is to show that when probabilistic explanation is merited at all, there is a single best probabilistic model for explaining the event; the probability that underlies this model is then the uniquely explanatory complex probability for the event. In this way, the "reference class" problem for explanation is solved.

There are as many kinds of complex probability as there are grounds for probabilistic explanation in deterministic systems. In what follows, I examine one particular kind of deterministic mechanism that produces Bernoulli patterns, which I call a *microconstant* mechanism. There may be others. But microconstant mechanisms are, I have argued in Strevens (2003b), responsible for a great proportion of the world's deterministically produced probabilistic patterns, including those found in the systems of statistical physics and of evolutionary biology.

10.22 Microconstant Processes

When a certain frequency of events is produced by what I will call a microconstant process, the best explanation of the frequency is, I will show, a probabilistic explanation. That is, when the optimizing procedure is applied to any deterministic model for the frequency, it finds a better model for the frequency that is probabilistic.

I will first, in this section, introduce the notion of a microconstant process and show how the microconstancy of such processes explains the probabilistic patterns they produce. The explanation is advanced entirely independently of the kairetic account. The next section then shows that the probabilistic explanation of the patterns is the best explanation. An important element is left to later (section 10.25): I will not yet assume a thorough-going determinism in the systems under discussion.

The Wheel of Fortune A classic and very simple example of a microconstant process is provided by a wheel of fortune. A wheel of fortune consists of a rotating disk, divided into different sections like a pie, and a fixed pointer. The wheel is spun, and allowed to come to rest. One particular section will be indicated by the pointer; the identity of this section determines the outcome of the process. A simple wheel might have its sections painted alternately red and black, in which case the outcome of the process is either *red* or *black*.

Such a wheel, as everyone knows, produces probabilistically patterned outcomes. The pattern in question is the Bernoulli pattern described above and typical of other simple gambling devices such as tossed coins, rolled dice, and roulette wheels. A series of outcomes displays long run order, in the form of a stable frequency—about one half *red* and one half *black* outcomes—but short term disorder. This Bernoulli patterning of the outcomes can be explained by two features of the process: a property of the wheel's mechanics, which I call *microconstancy*, and a probabilistic property of the initial conditions, which I call *macroperiodicity*. Microconstant processes take their name, obviously, from the first of these. I will characterize microconstancy and macroperiodicity, and then show how together, they explain the fact that the outcomes produced by a wheel of fortune are probabilistically patterned.

Microconstancy Suppose, for simplicity's sake, that the outcome of a spin on the wheel is determined by a single initial condition, say the speed with which the wheel is spun, designated v. Consider the function of v that is equal to one for values of v that produce a *red* outcome, and is equal to zero for those that produce a *black* outcome. Call it the *evolution function* for the wheel. Provided that the wheel is symmetrical in the way that such wheels are supposed to be, the evolution function will have a form similar to that pictured in figure 10.1.

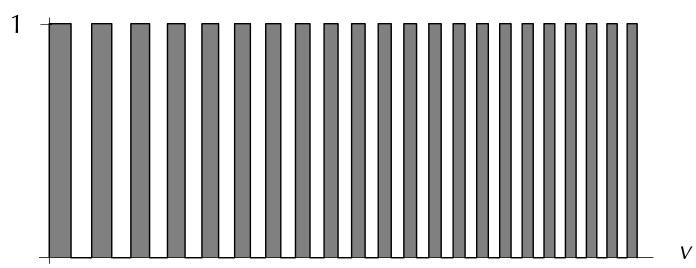

Figure 10.1: Evolution function for a simple wheel of fortune. Areas where the function is equal to one, signifying a value of v that produces *red*, are shaded gray.

The wheel's evolution function has two notable properties. First, it oscillates back and forth from zero to one very quickly. This reflects the fact that a very small change in the initial speed of the wheel can change the outcome from *red* to *black* and vice-versa. Second, there is a certain constancy to the oscillations: take any two neighboring *red* and *black* regions (gray and white in figure 10.1), and they bear the same ratio to one another anywhere along the graph—in this case, one to one.

I call an evolution function with these two properties *microconstant*. If a process has an evolution function for some outcome that is microconstant, the process, too, is called microconstant, relative to that outcome.

One more term is useful. I call the constant ratio of gray to white in an evolution function the *strike ratio* of the function. This is usually expressed as a fraction; for example, the strike ratio for *red* on the wheel of fortune is 0.5, because the proportion of the function that is gray (signifying the outcome *red*) for any neighboring pair of gray and white sections is one half.

Macroperiodicity Assume that there is a probability distribution over the initial conditions of the wheel of fortune, that is, over the wheel's initial spin speeds. (The question as to the basis of the distribution will be discussed below.) Macroperiodicity may be thought of as a geometrical property of the distribution's *density*. Let me briefly explain the nature of probability densities. The density for a distribution over a variable v is a function defined over v so that the probability that v takes on a value greater than x but less than y is equal to the area under the graph between x and y, as shown in figure 10.2. The density, then, is just a convenient way of representing the probabilities for

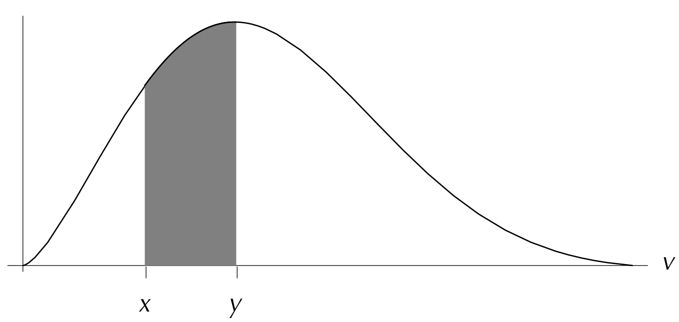

Figure 10.2: A density function over the initial condition v. The function is defined so that the probability of v taking on a value between x and y is equal to the area under the function between x and y, shaded in the figure.

all the events of the form *v is greater than x but less than y*. Perhaps the most familiar probability density is the normal curve.

A probability distribution is *macroperiodic* if the corresponding probability density is approximately uniform across almost all neighboring pairs of gray and white regions. Macroperiodicity is, then, a kind of smoothness: if a distribution is macroperiodic, the density changes, if at all, only gradually, so that neighboring initial conditions are about equally likely.

Strictly speaking, I am interested in what follows not in the probability distribution over the spin speed for a single trial, but the distribution over spin speeds for sets of trials. Such a *joint distribution* has a density with one argument for each trial in the set. If there are ten trials, the density has ten arguments, and gives the probability for every possible combination of ten spin speeds. I will require that these joint distributions be macroperiodic.

Normally, however, spin speeds for different trials are stochastically independent; given stochastic independence, if the probability distribution for a single spin speed is macroperiodic, then all joint distributions are macroperiodic. Thus I will remind the reader very seldom that it is the joint distribution that matters.

Explaining the Patterns The key to explaining the probabilistic patterns produced by a wheel of fortune is the following observation: a microconstant process with a macroperiodic initial condition distribution will tend to produce an outcome E with a frequency approximately equal to the strike ratio for E. The constraint placed on the initial condition distribution, note, is very weak; the frequency will be the same whatever the initial condition distribution, provided that it is smooth enough to count as macroperiodic.

Figure 10.3 tries to show graphically why this is the case for the wheel of fortune. Observe that the frequency of *red* will be equal to the probability that a given value of v produces a *red* outcome, which is equal to the shaded area under any of the densities in figure 10.3. When the evolution function for E is microconstant, this area will be roughly the same no matter what the density, provided that it is smooth.

More formal and more powerful is the following theorem: if a process has

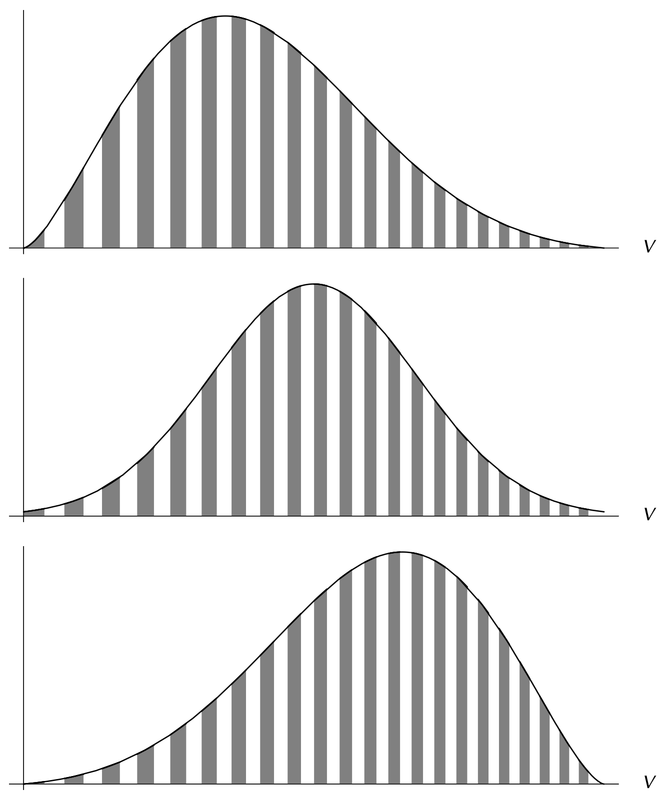

Figure 10.3: The shape of the initial condition density, if smooth, makes no difference to the frequency with which an outcome is produced: the shaded area of all three densities is one half.

a microconstant evolution function with strike ratio p for an outcome E, and the joint probability distribution over the process's initial conditions is macroperiodic, then the probability distribution over the outcomes produced by the process is a Bernoulli distribution that assigns events of type E a probability equal to p. Such a process will, then, tend to produce probabilistically patterned outcomes; more precisely, it will with a near one probability produce Bernoulli patterns with a frequency of E events equal to p. The longer the series of outcomes, the more certain they are to be probabilistically patterned. (An explanation using a roughly equivalent result was proposed by Poincaré (1896) and developed by Hopf (1934). The form of the explanation sketched here is worked out in detail in Strevens (2003b).)

Because the wheel of fortune is microconstant and the joint distribution over initial spin speeds for the wheel is (or so I assume) macroperiodic, the wheel produces Bernoulli patterns of *red* and *black*: in the short run, outcomes are completely uncorrelated, but in the long run, *red* and *black* are each almost certain to appear with a particular frequency, equal to the strike ratio of one half. The longer the run, the law of large numbers tells us, the closer to one is the probability that almost exactly one half the outcomes will be *red*, one half *black*.

A few remarks. First, the fact that, say, one thousand trials on a wheel of fortune produced about one half heads should be counted as an event, not a generalization, for the purpose of the study of explanation. We tend to say that the event of the frequency being such and such is realized by a series of events, but in saying this we are using *event* in a narrower sense. In this narrower sense, an event is something like a particular object's having a particular property, as opposed to a number of objects having a particular property. It is useful to be able to distinguish these notions; thus, I have been referring to the narrow kind of event as a *single event*. This should not be taken to imply that frequencies are not also kinds of events.

Second, microconstancy would be a mere curiosity if only systems like the wheel of fortune were microconstant. In Strevens (2003b) I make a case, as noted above, that ecosystems, the systems of statistical physics, and some other kinds of systems are microconstant with respect to many outcomes of interest. Furthermore, although microconstant mechanisms produce only Bernoulli patterns, systems with microconstant mechanisms at their heart can produce Bernoulli, Gaussian, Markovian, and other patterns, the complex systems just mentioned being examples.

Third, the explanation of the probabilistic patterns assumes, first, the existence of a probability distribution over the initial conditions, and second, that distribution's macroperiodicity. How plausible are these assumptions?

Taking them in reverse order, macroperiodicity is very plausible, because macroperiodic distributions over initial conditions are, as I have argued in Strevens (2003b), §2.5, the rule. The argument is long, however, and I will not reproduce it here. The prevalence of macroperiodicity is in any case not strictly speaking necessary for my purposes.

The existence of probability distributions over initial conditions is more problematic. This issue will be discussed in section 10.25. There is one point, however, that I want to make right away. Assuming a probabilistic explanation over the wheel of fortune's initial conditions does not imply that there is no deterministic explanation of the probabilistic patterns produced by the wheel; I give such an explanation in the next section. Thus the assumption will not prevent me from achieving one of my principal aims, to show that a probabilistic explanation is sometimes better than a competing deterministic explanation. What it does prevent me from showing, until I say more about the provenance of the initial condition distribution, is that probabilistic explanations can be the best explanations even in systems that are completely deterministic. This problem is resolved in section 10.25.

10.23 Microconstant Explanation

A wheel of fortune is spun 500 times. About one half of the outcomes are *red*. How, according to the kairetic account, is this approximate frequency to be explained?[2] I develop, first, a candidate deterministic explanation of the frequency, and second, a probabilistic explanation that is a formalization, using the language of probabilistic causal models, of the explanation sketched in the last section. The probabilistic explanation is shown to offer the better combination of accuracy and generality. Furthermore, the probabilistic model probabilistically generates the deterministic model. Thus the kairetic kernel corresponding to the deterministic model is the probabilistic explanation itself. Moral: sometimes, a less detailed probabilistic explanation of a frequency is better than a full deterministic explanation.

The Deterministic Model A deterministic model for the one-half frequency may be constructed as follows. Begin with a complete deterministic description of every spin of the wheel, from the moment that the wheel is set in motion to the moment that it comes to rest. The description of each spin will specify (a) the initial speed of the wheel, (b) everything relevant to the wheel's mechanics when spun at that speed, such as the composition of the wheel, its bearings, and the local air density, and (c) everything relevant to determining which states of the wheel determine which outcomes, in particular, the wheel's paint scheme and the pointer's position. (Note that the description omits mention of any probabilistic process that might have determined the initial speed; it is a deterministic model I am after here.)

These descriptions causally entail the outcomes of their particular spins; the concatenation of the 500 descriptions constitutes a setup for a causal model for the approximate one-half frequency of *red*. Because the model entails the fact that the frequency is roughly one half, it has an accuracy of one, the maximum possible. The model entails much more than this, of course: it entails

2. On explaining exact as opposed to approximate frequencies, see below.

the exact frequency for *red*, and indeed, the exact sequence of outcomes giving rise to this frequency.

Now apply the optimizing procedure to the model, but without allowing accuracy tradeoffs. The model will remain deterministic, but much detail can be discarded. For the description of any given spin, the initial speed need only be stated approximately, since an approximate speed is enough to determine whether the spin's outcome is *red* or *black*. For the same reason, the mechanics and paint scheme of the wheel need not be detailed down to the last molecule. Without these unnecessary elaborations, the description of each spin is just detailed enough to causally entail its outcome. The concatenation of the stripped down descriptions is the setup of a new deterministic model for the one-half frequency which has the greatest generality possible in a deterministic model. It is the deterministic champion.[3]

The Probabilistic Model My probabilistic model for the one-half frequency cites

1. The fact that the probability distribution over the initial spin speeds is macroperiodic, and

---

3. Here I ignore two competitors to the model I have proclaimed champion. The first competitor is a model that lists the different intervals of initial conditions that cause *red* outcomes and asserts merely that the actual initial conditions fell into one of the intervals on the list. I rule this model out of the running because it fails the cohesion test: by its own standards of individuation for causal factors, where factors correspond roughly to individual spin speeds, it uses a massive disjunction to describe the initial conditions section 3.6. Dismissing the model is not essential to my case, but it does simplify the exposition.

The second competitor is a deterministic analog of the probabilistic model described immediately below. It cites the microconstancy and strike ratio of the wheel's mechanism, and the fact that the distribution of the actual initial conditions for the series of 500 trials is macroperiodic (in a sense to be discussed further in section 10.25). These factors are together enough to *entail* a frequency of about one half. There is nothing wrong with this model; on the contrary, it is, in a certain sense, a part of the probabilistic model that I will favor. (The other part is a probabilistic part relating the probability distribution over the initial conditions to the distribution of the actual initial conditions. For further discussion that touches on this part/whole relation, see section 11.23.) Thus the second competitor and the probabilistic model stand or fall together. As such, the second competitor cannot be the deterministic champion to pit *against* the probabilistic model.

2. The physical facts about the wheel's mechanics, paint scheme, and so on that make it microconstant with strike ratio one half. Most important of these are the circular symmetry of the wheel's mechanics and the rotational symmetry of the red and black pattern of paint.

Together these entail a probability of near one for the frequency. The resulting model therefore has an accuracy of nearly, but not quite, one.

The kairetic account requires that the entailment of the near one probability for the explanandum, the frequency, be a probabilistic causal entailment. This means first, that all components of the model must be causal producers of the frequency, and second, that the entailment of the probability should mirror the way in which the frequency is causally produced. Let me show that my probabilistic model satisfies these requirements.

In a deterministic system, causal production is defined as before. Thus the rotational symmetry of the wheel, for example, qualifies as a causal producer for the frequency because it is physically realized by properties (the underlying physics of the wheel) that causally influence the events (the individual outcomes) that determine the frequency. The same is true of any other aspect of the wheel's physical makeup.[4]

Why is the initial condition distribution a causal producer? That depends on the physical foundation of the distribution. In a deterministic system, I assume, the distribution quantifies facts about the causal production of the actual initial conditions, so qualifies as a causal producer in the most straightforward way. (Various possible foundations for initial condition distributions are discussed in section 10.25.)

Now to show that the entailment of the probability in the model mirrors the way in which the causal factors in the model produce the explanandum,

4. Why is the paint scheme a causal producer? Because it figures in fundamental physics' causal story as to why an outcome is *red*. Causal producers do not have to exert a physical force on the outcomes they play a part in producing: consider facts about space, negative background conditions, and so on.

the frequency of *red*. Observe that the frequency is produced in two steps: first, the initial condition distribution produces a set of macroperiodically distributed actual initial conditions, and second, the evolution function takes these initial conditions and produces a set of outcomes with the given frequency. The application of the optimizing procedure shows that the only facts about the evolution function relevant to its producing the frequency are its microconstancy and strike ratio. The derivation of the probability, then, follows exactly the same path as the causal story about the production of the frequency.

Understanding probabilistic causal entailment in this way in deterministic systems puts a tight constraint, note, on the construction of a probabilistic model of a deterministically produced phenomenon: such a model, in entailing a probability, must also model the causal production of the corresponding frequency. Since complex probabilities exist just where probabilistic models succeed in explaining deterministic phenomena, this establishes a close relationship between the complex probability of an event and its frequency.

In section 9.72, I distinguished direct and indirect causal producers. The producers in my model—macroperiodicity, microconstancy, strike ratio—are, I propose, *direct* causal producers, because they are physical aspects of the system on which the explaining probability itself is founded, in the sense that the probability of one half for *red* can be seen as a quantification of the relevant properties of the initial condition distribution and the mechanism.

An indirect causal producer is anything that affects the direct producers, for example, a tampering with the mechanism of the wheel or an adjustment of the paint scheme.

The Optimal Model Now compare the deterministic and the probabilistic models for the one-half frequency of *red* outcomes. In accuracy, the deterministic model wins out, by a slender margin. For the purposes of assessing generality, the setups of both models can be divided into two parts: descriptions of the initial conditions, and descriptions of the wheel's mechanism, including the paint scheme, pointer position, and so on. In both cases, I will show, the probabilistic model is considerably more general.

First, the initial conditions. Where the deterministic model describes the 500 initial conditions of the 500 spins on the wheel that produce the frequency, the probabilistic model describes a property of a probability distribution over the initial conditions. I have declared that a description of initial conditions and a specification of a probability distribution over those same initial conditions ought to be considered roughly equal in generality. Perhaps this is overly generous to the deterministic model, since as the number of trials increases, the detail in the deterministic model's description of the initial conditions must increase, whereas a specification of the initial condition distribution will stay the same. Never mind.

If the probabilistic model specified the probability distribution over the initial conditions of the 500 spins, then, the two models would be about equally general in their characterization of the initial conditions leading to the one-half frequency. But the probabilistic model only specifies a single, simple property of the initial condition distribution, its macroperiodicity. This vastly underdetermines the shape of the probability distribution, thus is far more abstract than a specification of a complete distribution. With respect to the description of initial conditions, then, the probabilistic explanation of the one-half frequency of *red* outcomes is far more general than the deterministic explanation.

The same is true for the two models' descriptions of the wheel of fortune's mechanism. The deterministic model must describe the mechanism in enough detail to determine the outcome of a spin, given some particular initial spin speed. Factors such as the wheel's coefficient of friction, the exact number of red and black sections on the wheel, the wheel's position at the start of the spin, and the position of the pointer, will have to be specified at least approximately. The probabilistic model mentions only the symmetries of the wheel. To take an especially telling detail, it need say nothing about the position of the pointer, only that there is a pointer. This and the symmetries are enough to entail the wheel's microconstancy and the strike ratio for *red* of one half. The deterministic model's description of the mechanism is far more detailed, hence far less general, than that of the probabilistic model.

Overall, then, the probabilistic model is only slightly less accurate and is much more general than the deterministic model. Of the two, the probabilistic model offers a far better combination of accuracy and generality.

A side note: There is one kind of frequency that a microconstancy-driven model is not so good at explaining: an exact frequency. Suppose, for example, that of 500 spins of the wheel of fortune, exactly 250 yield *red* outcomes. What explains the fact that the frequency is *exactly* one half? The probability assigned to this event by the Bernoulli distribution is quite low: about 0.04. Thus the microconstant model—and indeed any probabilistic model that cites a one half probability for *red*—has very low accuracy. Whether this is enough to undo the generality advantage it enjoys over the deterministic model is hard to say, but if so, then it is the deterministic model, or something like it, that best explains the exact frequency. (This is rather a point of principle, since scientific explanations are rarely explanations of exact frequencies.)

Can the probabilistic model be further refined? There are two ways this might be done. First, the accuracy of the model might be improved by adding a little information about the initial conditions or the wheel's mechanism, creating a slightly less general model. But adding even quite vast quantities of information about either the initial conditions, the mechanism, or both, does not improve accuracy at all. Only at the point at which the model becomes deterministic—at which it entails a frequency for *red* of one half—does accuracy improve, and then only slightly.

Second, the generality of model might be improved by removing information about the initial condition distribution or the mechanics of the wheel. Taking information away, however, almost immediately reduces the accuracy to zero. The probabilistic explanation cannot be improved.

Perhaps there is some completely different approach to abstracting away from the deterministic model that produces something better than my probabilistic model. But I cannot see how. What is so striking about the probabilistic model is that it is both extremely general—it contains only a few pieces of information—and extremely accurate.

The probabilistic model for the one-half frequency is not only better than the deterministic model; it probabilistically generates the deterministic model. This is because the details of the wheel's mechanism laid out in the deterministic model entail the microconstancy and the one-half strike ratio of the mechanism, and the initial condition distribution both probabilistically generates the initial condition description and entails the macroperiodicity of the distribution. Thus when the optimizing procedure is applied to the deterministic model, it yields the probabilistic model. The probabilistic model is the kernel corresponding to the deterministic model.

It follows that the only direct causal producers explanatorily relevant to the production of the one-half frequency are the facts stated in the probabilistic model: the macroperiodicity of the initial condition distribution, the circular symmetry of the wheel's mechanics, and the rotational symmetry of the paint scheme.[5] The factors that appear in the best deterministic model but not in the probabilistic model are, as noted in section 10.1 explanatorily irrelevant difference-makers: they make a difference to the explanandum, but not enough of a difference to be explanatory.

The kairetic account's judgment that the probabilistic explanation of the frequency of *red* is optimal captures our explanatory intuitions exactly. First, the high level facts specified in the probabilistic explanation are indeed essential to understanding the frequency. You do not truly understand why one half of the outcomes were *red* until you see the role that the symmetry of the

5. Plus of course a few other facts about the operation of the wheel to obvious for me to spell out in the main text: that it is spun with a speed determined by the initial condition distribution, that the outcome is determined by a pointer and the paint scheme, and so on.

wheel and paint scheme play in creating microconstancy with strike ratio one half. Second, the facts about the initial conditions and the wheel's mechanism not specified in the probabilistic explanation are of no explanatory interest. Piling on the details about which spins went where when does not illuminate the one-half frequency at all. It explains the individual outcomes, but it does not explain the fact that those individual outcomes are one half *red*. It is not too much of an exaggeration to say that Laplace's demon, however well it may understand individual *red* outcomes on the wheel, has no clue why the long run frequency of *red* is one half.