Dynamic Probability and the Problem of Initial Conditions

Michael Strevens

Forthcoming, *Synthese*

Abstract

Dynamic approaches to understanding probability in the non-fundamental sciences turn on certain properties of physical processes that are apt to produce "probabilistically patterned" outcomes. The dynamic properties on their own, however, seem not quite sufficient to explain the patterns; in addition, some sort of assumption about initial conditions must be made, an assumption that itself typically takes a probabilistic form. How should such a posit be understood? That is the problem of initial conditions. Reichenbach, in his doctoral dissertation, floated a Kantian solution to the problem. In this paper I provide a Reichenbachian alternative.

Acknowledgments

My thanks to audiences at the Complutense University in Madrid, the "All Things Reichenbach" conference at the Munich Center for Mathematical Philosophy, and two anonymous referees, for insightful comments and advice.

1. Introduction

Chance, if our scientific picture of the world is to be taken at face value, is everywhere. That is, a tremendous range and variety of scientific models, across physics, biology, and the behavioral, medical, and social sciences, purport to predict and to explain phenomena by imposing probability distributions over things seen and unseen. These probability distributions are to all appearances "out there", supervising the course of events, making things happen.

Is that the way it really is? Even if we understand quantum mechanics so as to impute stochasticity to the fundamental laws of nature (by way of a collapse theory, for example, such as the GRW interpretation), the resulting probability distributions do not seem sufficient to provide an underpinning for the great majority of successful statistical models in the higher-level sciences.

Consider, for example, the humble coin toss. The dynamics of the toss are deterministic (or so close as makes no difference), in the sense that the initial conditions of a toss—in particular, the rotational and translational motions of the coin—fix whether the coin lands heads or tails. Yet despite this determinism, a model that ascribes to heads a probability of one-half has great predictive and explanatory merits. We would like to say that the one-half probability captures, if not the whole truth about the coin's dynamics, then something of considerable importance. That "something" is our physical probability (or "chance"). It is not the quantum probability of heads given the initial conditions, which is zero or one (or near enough). So what could it be?

Perhaps it is nothing, or at any rate, nothing over and above the existence of certain general patterns in the tosses' outcomes, such as the one-half frequency of heads. Those are the views of the subjectivist and the actual frequentist, respectively (and very roughly). Perhaps it is a manifestation of more distal quantum probabilities—if they impose a probability distribution over the tosses' initial conditions, say, that in turn imposes a one-half probability on heads.[1]

Or perhaps—and this is the dynamical approach—it is some property of the tossed coin's physics, of the mechanical process by which a toss's initial conditions lead to its outcome, heads or tails. The dynamical approach, pioneered by the German psychologist Johannes von Kries in 1886, has tempted

1. Albrecht and Phillips (2014) make such a proposal. Strevens (2003) provides a general framework for thinking about such scenarios, under the guise of "simplex probability".

numerous deep thinkers about probability since, among them Henri Poincaré and Hans Reichenbach, and more recently philosophers such as Marshall Abrams, Wayne Myrvold, and myself.[2] Yet it faces considerable empirical and conceptual challenges.

Perhaps foremost among these is the question of the role of initial conditions in the dynamical approach. It seems that an initial-condition distribution of some sort must play a role in the dynamical story, yet what sort and what role? And does the invocation of what appears to be an antecedent probability distribution render the entire exercise pointless, by assuming the existence of physical probability where it was supposed to be explained? In what follows, I set out to sketch some answers to these questions.

My strategy will be indirect. Rather than plunging into semantics or metaphysics, I inquire into the explanatory foundations of statistical models in the higher-level sciences—that is, I ask what underlying properties of the modeled systems furnish the models with their explanatory power. I then propose to identify physical probability or chance with these explanatory structures.

This assimilation of explanatory and metaphysical questions is, I believe, a commitment that is characteristic of those taking the dynamical approach to understanding physical probability. I will not attempt to defend it here; such a defense would require an excursion into meta-metaphysics that is best left to another time and place. Some readers may therefore remain skeptical. I hope, however, that even they will find in this paper an illuminating blueprint for understanding the explanatory underpinnings of many high-level stochastic

2. See von Kries (1886), Poincaré (1896), Reichenbach (2008), Strevens (2003), Abrams (2012), and Myrvold (2012, 2021). Other mathematicians who have developed the idea of dynamic probability along the lines followed in this paper include Hopf (1934), Keller (1986), and Engel (1992). Much of the modern ergodic approach to the foundations of statistical mechanics might be understood as mining the same vein (Sklar 1993). Further work by philosophers developing or commenting on dynamic probability includes Rosenthal (2010), Butterfield (2011), Strevens (2011), Filomeno (2019), Suárez (2020), and de Canson (forthcoming). A historical treatment may be found in von Plato (1983).

models.

2. Dynamical Probability

A toy system that is especially useful for introducing the idea of dynamical probability because of the simplicity of its physics is a simple wheel of fortune. The wheel, painted in alternating red and black sections like a roulette wheel (but without the zeroes) is set spinning on its axis; when it comes to rest, a stationary pointer indicates some particular section of the wheel, red or black. That is the outcome of a spin on the wheel. With just a little idealization, we can think of the outcome as being entirely determined by the wheel's initial spin speed along with various fixed physical parameters: the spinning mechanism's coefficient of friction, the paint scheme, the position of the wheel before it is spun, the position of the pointer, and so on. (Assume that the wheel is returned to the same starting position before every trial.) A given initial spin speed, then, will deliver the same outcome every time. The wheel is fully deterministic.

Now consider a function that, for a given wheel of fortune, maps each possible initial spin speed to its outcome. For convenience, let's say that the function takes the value 1 for speeds that yield *red* and 0 for speeds that yield *black*. I call this the wheel's *evolution function* for *red*. Graph the evolution function with the spin speeds laid out along the horizontal axis, and you will get something like the plot shown in figure 1. I have shaded the area under the function, so that the gray areas span just those values of the initial spin speed that yield *red*.

There is something rather suggestive about the evolution function, something that at first blush promises to explain the distinctive pattern of outcomes produced by devices like the wheel of fortune.

Let me say a few words about that pattern. As every savvy gambler knows, it has two important features. On the one hand, in the long run, *red* tends to appear with a stable frequency of about one-half. The frequency may not

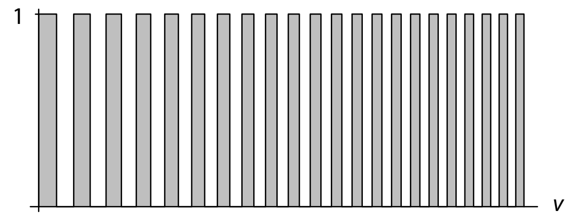

Figure 1: Evolution function for the outcome *red*, for a simple wheel of fortune set in motion with initial spin speed v

be a sure thing, but it is close enough: on a wheel that is spun sufficiently many times, you can with great confidence predict that about one-half of all outcomes will be red and one-half black. On the other hand, there is no scope for prediction in the short run. The outcomes are (again, usually if not always) thoroughly scrambled; there is no way to determine, given the pattern of outcomes so far, what the next outcome is likely to be.

These patterns are characteristic of any Bernoulli process, that is, any two-outcome process in which the outcomes occur with fixed probabilities (not necessarily one-half) and in which they are statistically independent. The characteristics of these *Bernoulli patterns*, then, are that the outcomes occur in the long run with a stable frequency, but in the short run in an otherwise totally disordered sequence. The patterns of heads and tails produced by tosses of a coin (whether fair or biased), the patterns of sixes and non-sixes on a rolled die, the patterns of *red* and non-red (including zeroes) on a roulette wheel: all have the characteristic Bernoulli aspect.[3]

Why should these deterministic (or near-deterministic) processes produce Bernoulli patterns? Look at figure 1. It seems to supply the answer.

The short-term disorder characteristic of a Bernoulli pattern is the result of the rapid alternation of the function between zero and one, as you travel along

3. The notion of a Bernoulli pattern is easily extended to processes with many outcomes, such as the six outcomes normally noted on a die roll. But two outcomes will be quite sufficient for my purposes.

the horizontal initial-condition axis. Small differences in initial conditions are enough to reverse the outcome of a trial (a spin, a toss, a roll). Outcomes are therefore very difficult to predict, even given considerable information about the distribution of initial spin speeds.

The long-term order, meanwhile, is due to the same rapidity of the alternation combined with its regularity—the fact that it swings back and forth with a fixed rhythm, or in other words, that the ratio of neighboring gray and white regions in the evolution function is constant throughout the range of possible initial conditions. As a result of this property, which I call *microconstancy*, almost any reasonably smooth distribution of initial conditions will produce outcomes with the same proportions. In the case of the wheel of fortune, that means equal numbers of *red* and *black*. More generally, when the constant ratio of gray to white, which I call the *strike ratio*, differs from one-half, almost any reasonably smooth distribution will produce "gray" outcomes in proportion to the strike ratio. (A smooth distribution is one that does not vary too much over narrow intervals; much more on this shortly.)

This fact can quickly and easily be grasped by visual means. Suppose that the distribution of initial spin speeds on a long series of spins of the wheel of fortune can be captured, at least approximately, by a density function of the sort shown in figure 2. Such a function is to be interpreted as follows: the proportion of initial spin speeds lying between two values is equal to the area under the function between those values, as a proportion of the whole.[4] In the figure, the proportion of spin speeds lying between x and y is therefore roughly equal to the proportion of the area under the function that is shaded.

We can then superimpose the density for a given long series of spins on the evolution function for the wheel, as shown in figure 3. The proportion of *red* outcomes in the series will equal the proportion of the density that is

4. A rough correspondence between the density and the actual distribution of spin speeds is good enough for my expository purposes in this section; I will treat the issue more carefully in section 3.4.

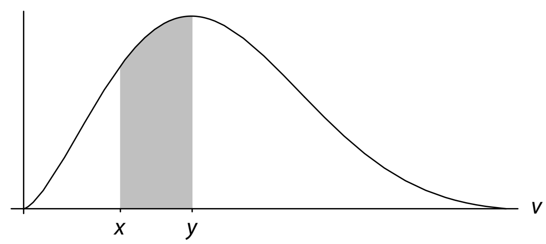

Figure 2: The shaded area is proportional to the percentage of spin speeds (in some series of spins) whose values lie between x and y

shaded gray. As you can see, many different densities will yield a frequency of *red* approximately equal to the strike ratio, that is, approximately equal (in the case of the wheel) to one-half. Indeed, as I have said, any reasonably smooth density—any density that is close to uniform over short stretches—will do so.

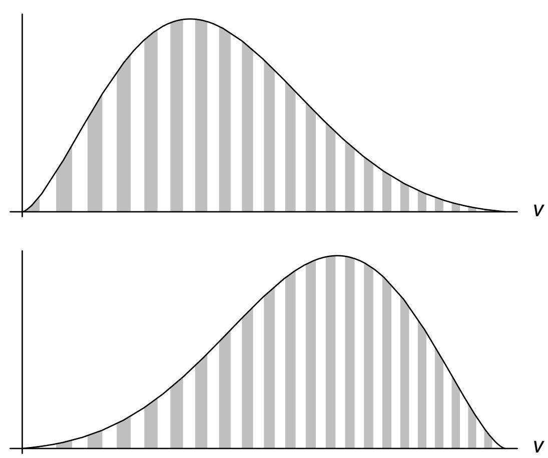

Figure 3: Different distributions of initial spin speeds yield *red* with equal frequency

It is hard to resist the conclusion that somewhere in the vicinity of microconstancy lies the explanation of the fact that wheels of fortune and like devices, constructed of different materials and operated by numerous different croupiers across varied cultures and times, have all produced and continue to produce outcomes with fixed, stable long-term frequencies, now seen to be equal to the outcomes' strike ratios. Likewise, the patterns characteristic of other paradigmatic probability distributions, such as Gaussian distributions, Poisson distributions, Markov chains, and so on—I call them, collectively, the *probabilistic patterns*—might be elegantly explained by their own proprietary dynamic processes.

It's no surprise, then, that since von Kries first explored dynamic explanations of the Bernoulli and other probabilistic patterns, the project has been revived and newly touted many times (see note 2). Evidently, the notion that the Bernoulli patterns produced by certain systems can be explained by a distinctive feature of their underlying dynamics—microconstancy—is an enduring one. Yet it has never quite broken into the mainstream. There are, I think, two deep reasons for this.

First, the microconstancy of even a relatively simple gambling setup such as a rolled die is more a surmise than a known fact.[5] And in any case, the canonical gambling devices are of limited interest to those outside the casino business. Statistical models in science explain thermodynamics, evolution by natural selection, and much more. If the dynamical approach to probability is to take a star turn in understanding the basis of successful stochastic modeling, we must have some reason to think that the relevant underlying physical, biological, and other processes have microconstancy or some similar property. To make a case for this conclusion is no small undertaking. I do my best in Strevens (2003), but it is a speculative enterprise.

Second, a dynamical property such as microconstancy cannot explain a probabilistic pattern unaided. The story must be supplemented by some assumption about initial conditions. It is quite unclear, as I have said, what that assumption ought to be, and indeed whether making such an assumption voids the interest of the dynamical project, positing probability where it was

5. Strevens (2013), chap. 6.

supposed to have been explained. The remainder of this paper will be focused on this problem of initial conditions.

A third question lingers in the background, of how to proceed from an explanation of the Bernoulli patterns to a thesis about the nature of chance. I have proposed that we identify a physical probability distribution posited by a statistical model with the features of the modeled system that explain the patterns of outcomes predicted by the distribution. We should identify the one-half probability for *red* on a simple wheel of fortune, then, with the aspects of the wheel's physics that undergird its microconstant dynamics (and determine the strike ratio for *red*) together with some property of the initial-condition distribution. For the greater part of this paper, however, I will put the metaphysics to one side and focus exclusively on the dynamical explanation of the patterns.

3. The Problem of Initial Conditions

3.1 Making Sense of Distribution Talk

Suppose that the distribution over the initial spin speeds on the wheel of fortune were captured by the density shown in figure 4. Then the frequency

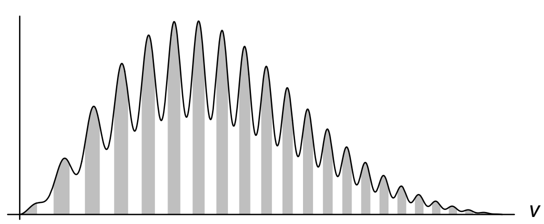

Figure 4: A distribution of initial spin speeds that is not at all smooth, disproportionately favoring speeds that produce *red*

with which the wheel produced *red* outcomes would be, very likely, something quite a bit larger than one-half (since the shaded area under the density, corresponding to spin speeds that lead to *red*, is considerably more than one-half of the total area). To explain an actual frequency for *red* of one-half, then, it seems that it is not sufficient to point to the wheel's dynamic properties alone. We also need, at the very least, to posit that the initial-condition distribution is *not* of the sort shown in figure 4, or anything like it. Using the informal characterization I introduced above, we need to posit that it is "reasonably smooth".

Such a postulate raises two questions. First, what property, exactly, are we attributing to the distribution? What is this "smoothness"? Second, what distribution, exactly, are we attributing it to? If a full-blown probability distribution, then where does that come from? If not, what?

The first question is by far the easier to answer. A device's dynamics is microconstant (relative to a certain outcome) just in case its space of initial conditions can be divided into small, contiguous sets, within each of which approximately the same proportion of conditions produce the outcome. (That proportion is the strike ratio for the outcome.) A conception of "smoothness" that naturally dovetails with microconstancy so as to produce outcomes with a frequency or probability (or whatever the initial-condition distribution measures) equal to the strike ratio is this: approximate uniformity over these same small sets. Note that smoothness is therefore relative to an outcome and a dynamics; what is smooth for one device or outcome might be unacceptably turbulent for another.[6]

The finer aspects of this conception of smoothness can be tweaked in various ways, but I will not concern myself with the details. I will simply label this property (or family of closely related properties) smoothness, and note that any smooth distribution over the initial conditions of a microconstant dynamics for a given outcome will attribute a probability (or whatever the

6. For a full explanation of the Bernoulli patterns over a series of trials, we must derive not only the long-term order characteristic of the patterns, that is, the stable long-term frequency, but also the short-term disorder, or statistical independence, of individual outcomes. I let the question of independence lapse for now; I will take it up again in section 3.4.

distribution measures) to the outcome equal to its strike ratio.[7]

Which naturally leads to the second question: what does the distribution quantify? Or rather—the correct way to pose the question given my explanatory agenda—what ought it to quantify in order that, when partnered with microconstancy, it will yield a bona fide scientific explanation of the Bernoulli patterns?

In sketching the explanation of the wheel of fortune's Bernoulli-patterned outcomes above, I talked as though the initial-condition distribution represented the statistical profile of the actual spin speeds involved in the series whose pattern was to be explained. That is an enticingly straightforward interpretation of the distribution, but a frequentist construal was not the strategy pursued by the founders of the project: von Kries, Poincaré, and indeed, in his dissertation work, Reichenbach. Let me take a look at their various approaches.

3.2 *Doing Without an Interpretation*

It would be a neat trick, and much metaphysicking might be avoided, if it were possible to explain the patterns produced by the wheel of fortune without committing to any particular interpretation of the initial-condition distribution at all.

The idea that comes closest to pulling off this trick is Rosenthal's (2010) proposal that microconstancy alone is sufficient for the existence of a physical probability equal to the relevant strike ratio. Such a probability depends only on the dynamics, and not at all on some distribution of initial conditions.[8]

What if there is a distribution of initial conditions, however, and what if it looks like figure 4? To avoid ascribing a probability equal to the strike ratio in

7. In earlier work, I have called the property of smoothness "macroperiodicity" and "microequiprobability". In opting for the simpler "smoothness" in this paper, I trust that the reader will go along with my use of a common word as a term of art.

8. A similar line of thought is pursued in unpublished work by Wolfgang Pietsch.

such cases, Rosenthal imposes certain dynamical conditions on the processes that produce the initial conditions. That takes the story rather closer to my own proposal in this paper, while continuing to avoid any distribution talk.[9]

Yet a voice in my head tells me that it is impossible to account for probabilistic patterns, such as the Bernoulli patterns, without saying something about the distribution of initial conditions. To explain the patterns is to succinctly summarize their causal history, and the initial conditions are an essential part of that history. They cannot be omitted from the explanatory narrative entirely.

Poincaré, in his development of the dynamical approach, takes the initial-condition distribution to be relevant but thinks that we can explain the Bernoulli patterns "from the objective point of view" while saying very little about it.[10,11] It is enough that we have a secure basis for believing the distribution, whatever it is, to be smooth.[12] Whence this knowledge? Poincaré writes that he is "naturally induced" to suppose smoothness (p. 201). This line of thought should be distinguished, I think, from an appeal to a principle of indifference to warrant a smooth subjective probability distribution over the initial con-

9. I do not have space to discuss Rosenthal's approach any further here, but it is worthy of serious attention. For a presentation that provides, in addition, a very nice overview of various attempts to deal with initial conditions, see Rosenthal (2016).

10. Poincaré (1905), p. 202. Unless otherwise noted, the quotations in this paper are from Poincaré (1905).

11. Poincaré sketched his treatment in philosophically informal terms in a number of works. In some of these sketches he focuses on the distribution not of initial spin speeds but of the total angle turned by the wheel, which is determined by the initial speed and the wheel's physics. The presentation in Poincaré (1914), however, focuses on the initial speeds.

12. Poincaré experimented with several different notions of smoothness, all close to the informal characterization used in this paper. There is a certain tension between his smoothness assumptions and his claim that "As for [the initial-condition density], I can choose it in an entirely arbitrary manner" (p. 201). Not every density is smooth in Poincaré's sense(s), and so the density cannot in fact be chosen arbitrarily. Yet in the interpretation and development of Poincaré's ideas, subsequent thinkers—including, as we will see, Hans Reichenbach—have tended to emphasize the importance of arbitrariness, to the point where the dynamical approach has become known as "the method of arbitrary functions". In my view that is rather unfortunate, as any sound approach to explaining the probabilistic patterns must be based on smooth, and hence not entirely arbitrary, densities.

ditions. Poincaré's talk of an objective explanation suggests to me that the distribution is itself to be understood as an objective entity, concerning which his "natural inclinations" give him reliable knowledge.

Even if this somewhat mysterious source of knowledge is allowed, however, we philosophers—we interpreters of the world—would like to know what, precisely, it is that we are assuming to be smooth, and thus what, precisely, we are using to stage our explanation of the Bernoulli patterns.

A more philosophically concerted effort to evade engagement with a worldly initial-condition distribution can be found in recent literature on "typicality".

The idea, presented with reference to the dynamics of a Galton board by Maudlin (2007), is as follows. The pattern to be explained is not the ultimate arrangement of balls at the bottom of the board (a Gaussian distribution), but rather the pattern of bounces as a ball travels downward to its final resting place. Hitting a pin on the board, a ball can go either left or right. Record a sequence of such outcomes and you will find the same Bernoulli pattern you see in the wheel of fortune (with "left" and "right" for *red* and *black*). The dynamics generating these patterns is (at least for pins sufficiently far down the board) microconstant. Thus, as with the wheel of fortune, we could account for the pattern by pointing to the dynamics of the board and the smoothness of the distribution of the initial conditions, that is, the position and velocity of the balls as they are dropped into the board.

Maudlin offers an explanatory strategy that makes no reference to a physically real initial-condition distribution. It is sufficient to explain the Bernoulli patterns, he suggests, that almost any distribution of initial conditions for the insertion of balls into the machine will, given the dynamics, produce Bernoulli-patterned bounces. The dynamic behavior that produces the Bernoulli patterns is in this sense "typical", where the typicality of a dynamics is a mathematical, not a statistical, property that depends only on whether the sets of initial conditions that produce it constitutes a large proportion of all possible sets of initial conditions (assessed by any "reasonable" measure). Maudlin is quite clear that we are not using the property of typicality to infer something about a certain, explanatorily relevant distribution of initial conditions, as does Poincaré; rather, the fact of typicality itself does the explanatory work that we would otherwise assign to such an initial-condition distribution.

The viability of this strategy hinges, then, on the validity of appeals to typicality for explanatory purposes. Such appeals have yet to find a place in the mainstream literature on scientific explanation, although their vindication is currently an active area of philosophical research (see Lazarovici and Reichert (2015), Maudlin (2020), and Wilhelm (in press) among other recent work). I have my suspicions, but the debate is too complex to enter into in this paper. Let me put typicality aside, then, and move on.

Still another way to explain the Bernoulli patterns without making substantive assumptions about a mind-independent initial-condition distribution is to opt for mind dependence, interpreting the initial-condition distribution as purely epistemic—that is, as a distribution of subjective probabilities or something similar. Its smoothness might be a matter of logic or reasonableness, or of conditionalizing on empirical data (perhaps observations of the Bernoulli patterns themselves), but either way, the distribution is not itself a feature of the physical world. This is the kind of approach advocated by Myrvold (2021).

Some sort of epistemic conception of explanation, such as Hempel's (1965) expectability account, might then endow the distribution with explanatory power: we ought to expect the distribution of initial conditions to be smooth, so we ought to expect the outcomes to manifest the Bernoulli pattern. In that expectation, says Hempel, lies explanation.[13]

13. Frigg (2011, 91) outlines such an appeal to the expectability conception of explanation, though he is himself reluctant to endorse it. Myrvold cites Frigg, but it is difficult to tell whether he himself endorses the expectability approach. He does, however, argue that a smooth subjective probability distribution can explain why a thinker aware of microconstancy would expect to see Bernoulli patterns (or rather, the equivalent for the case of statistical

That will be good enough for some thinkers, but it is not good enough for me. The weight of philosophical opinion among philosophers of explanation is rightly against Hempel's expectability doctrine; explanation is not even partly epistemic, not even sometimes. To explain Bernoulli patterns in a series of trials, we need to cite some causally relevant fact about the trials' initial conditions.

3.3 Reichenbach's Constructivism

A distinctive attempt to avoid having to find an initial-condition distribution in the outside world is set out in Hans Reichenbach's doctoral thesis. Reichenbach adopted a neo-Kantian framework: "In its scientific sense experience is a representation of reality that combines given contents of perception through fixed [synthetic] a priori forms of classification" (Reichenbach 2008, 137).[14] Chief among these "forms of classification" is the "principle of lawful connection", which gives our "scientific experience" (or knowledge) its causal character. The causal structure of the world as we encounter it through experience is, then, as Kant himself held, an imposition of the mind.[15]

The principle of lawful connection "only describes the general form that specific experience fills with specific content" (131); what is a priori, then, is that the world's goings-on are orchestrated in accordance with causal laws

mechanics). I think that is correct, though I interpret such an explanation as straightforwardly causal: psychological facts (subjective probabilities, knowledge of or belief in microconstancy) cause the phenomenon to be explained, namely, a certain expectation (or in Myrvold's setup, the alleviation of puzzlement—"Oh, I see, I should have expected that all along").

14. Reichenbach refers to the a priori framework as "synthetic" in many other passages, though not this one.

15. In place of "experience in a scientific sense", Reichenbach often refers to physical knowledge: "Physical knowledge consists in the coordination of mathematical equations with particular objects of empirical intuition" (123). He appears to be halfway between the Kantian picture in which the principle of causality plays a part in constituting spatially and temporally structured experience from unstructured sensory input, and the empiricist picture in which the principle of causality plays a part in constituting knowledge from spatially and temporally structured percepts. Nevertheless, the rules to which the constitution conforms are explicitly said have the characteristic Kantian epistemic status: they are synthetic a priori.

that impose upon events a suite of regularities. The mathematical form and relata of the laws and regularities must be determined a posteriori, ideally by scientific inquiry. We posit those laws that best explain, or predict, the patterns we find in our percepts.

The lawful connection principle is not sufficient, however, for this task. (Here, I take it, Reichenbach follows Kant in supposing that the principle supplies only deterministic laws and causal connections.) The main reason is that in practice, we can derive predictions from a causal law or theory only if we assume that any interfering factors have a small or negligible impact on the course of events. (We would not need such an assumption if we had complete knowledge of the relevant boundary conditions and the ability to compute their consequences—hence "in practice".) The canonical form for these assumptions—about errors in the measuring apparatus, fluctuations in the causal background, and so on—is stochastic. We apply a causal theory, then, in the following way: we take account of errors and fluctuations; we determine that the probability of a serious perturbation is low; we then go ahead and use the theory to make a prediction. If the prediction is vindicated, the law or theory is confirmed. In this way, the principle of lawful connection becomes useful to us as an organizer of our knowledge about the world.

If we were to lack the ability to make stochastic assumptions about interference, however, prediction, confirmation, and thus the causal organization of knowledge would be impossible. Consequently, that ability is a precondition for scientific knowledge:

> Hence we conclude that the principle of lawful connection ... is insufficient for the mathematical representation of reality. A further principle has to be added ... the principle of lawful distribution (127).

As the first principle gives us the ability to represent deterministic causal laws, so the second principle gives us the ability to represent probabilistic distributions over initial and background conditions.

Such a probability distribution might be sufficient in itself to represent a pattern of perturbations. But it might also be put together with certain causal representations, such as that of the microconstancy of an ensuing process, to create a dynamical probability distribution, a partly causal form of representation that captures Bernoulli and other probabilistic patterns.[16] This is a distinct way, then, in which the principle of lawful distribution can work with the principle of lawful connection (which supplies the apparatus for dynamical representation) to organize our experience—in the case of the wheel of fortune, to organize our perception of Bernoulli patterns using the inherently dynamic concept of a microconstant probability distribution. Hence Reichenbach's interest in dynamic probability, which is developed at some length in the thesis.

In the same way, I presume, that every event falls under the principle of lawful connection—the mind provides sufficient connective scaffolding that every event has a cause—so every event falls under the principle of lawful distribution, which is to say, the mind provides sufficient distributive scaffolding that every event is subsumed under one (or more) probability distributions. We can know a priori, then, that the initial spins of a wheel of fortune have such a distribution.

All we need in order to derive (cognize?) the dynamical probability is the posit that the distribution is smooth.[17] Reichenbach appears to believe that not only the existence but also the smoothness of a distribution over initial conditions can be known a priori. One half of the reason is that he takes the lawful distribution principle to guarantee not merely a probability distribution but a probability density—that is, a probability distribution that is absolutely

16. For mixes of distribution and dynamics that might in this way represent the distribution of a measurement device's errors, see Strevens (2013), chap. 11.

17. Smoothness is not, of course, required for the lawful distribution principle to represent an error distribution whose parameters are discovered empirically. In these cases, "the principle of distribution only describes the general form that specific experience fills with specific content" (131).

continuous (relative to the relevant measure) (p. 129). The other half, I surmise, is that he is encouraged by various idealized dynamical models to suppose that absolute continuity is smoothness enough to derive the probabilistic patterns.[18] As an explanatory move, that seems rather questionable: the patterns to be accounted for exist in this non-ideal world of ours, in which a stronger formulation of smoothness is needed to get the relevant explanations up and running. But never mind.

Reichenbach soon left the Kantian worldview behind. Laws of nature and physical probability distributions are, he continued to believe, instruments for the representation of perceptual patterns, but they ought not to be understood as contributing any substantive content to such representations. They are mere cataloging devices, advanced filing systems for the efficient acquisition and retention of information about what is perceived. Reichenbach had become an empiricist.

3.4 Frequentism

For an empiricist, a statistical model is a tool for representing patterns of outcomes. The most direct implementation of this precept is the actual frequentist's: a model imposing a physical probability distribution over a given set of outcomes does nothing more nor less, says the frequentist, than assert that outcomes of that type occur in the pattern corresponding to the distribution. A model that puts a fifty-fifty Bernoulli distribution over *red* and *black* on the wheel of fortune, for example, asserts that *red* outcomes occur in the long run with a frequency of about one-half, while being in every other respect entirely disordered.[19]

18. In so doing, Reichenbach would be following the program implicit in Poincaré's remark that the initial condition density can be "chosen arbitrarily" (see note 12).

19. In a universe with infinitely many spins of the right sort, the long run might comprise infinitely many outcomes, in which case the long-run frequency is to be understood as a limiting frequency.

More must be said. What is the scope of the long run? How ought the disorder to be characterized? Reichenbach gives answers to these questions and many others in his great handbook of frequentism, *The Theory of Probability* (Reichenbach 1949). Let me put aside these issues in their most general form, however, to focus on the special application of frequentist thinking that is germane to the present discussion, namely, the interpretation of a distribution over the initial conditions of a microconstant process, and in particular, the meaning of smoothness for such a distribution.

The frequentist idea is of course to interpret the smoothness claim as a statistical claim about the distribution of some set of actual initial conditions; in the case of the wheel of fortune, then, some set of actual spin speeds. But what claim? And which set?

First, the set. Given my explanatory agenda, it seems that the set ought to contain all and only the speeds of the spins that produce the Bernoulli pattern in question. It is these speeds, after all, that are causally responsible for the outcomes, and thus for the patterns to which the outcomes conform.

And second, the claim. Though there may be many such spins, their number will be finite. We need to give some sense to the assertion that a set of finitely many spins is "reasonably smoothly distributed". Explanation should be our guide. What property of a finite set of spin speeds, together with the microconstancy of the wheel of fortune's dynamics, will explain its outcomes being Bernoulli patterned?

It is clear enough in outline what the answer should be. By the nature of microconstancy, the range of initial spin speeds can be divided into many small intervals, in each of which the same proportion of speeds leads to *red* (equal to the strike ratio of one-half). Take such an interval and divide it into not-too-small subintervals such that (a) each subinterval is of equal size, and (b) the speeds spanned by any given subinterval all lead to the same outcome (either *red* or *black*). In terms of the gray and white of figure 1, we want each of our equally-sized subintervals to be either entirely gray or entirely white. (It is sufficient that almost all our subintervals satisfy this condition; a small number of violations are permitted.) Then say that the initial-condition distribution is smooth just in case, within any such interval, the number of actual spin speeds falling into any of the subintervals is roughly the same. (Again, a small number of exceptions may be tolerated.) It follows that within any such interval, the proportion of actual spin speeds leading to *red* is approximately equal to the proportion of possible spin speeds leading to *red*—in graphical terms, the proportion of the interval that is shaded gray. Together with microconstancy, this "smoothness" yields a causal explanation of the fact that, in the series of spins in question, *red* occurs with a frequency equal to its strike ratio of one-half.[20]

Using the same sort of technique, we can define a probability distribution over the initial spin speeds, if they are smoothly spread. Use a bar graph to represent the number of actual spin speeds falling into each subinterval. Then join the tips of the bars with a line, skipping those subintervals, if any, that violate the uniformity requirement (i.e., that deviate to a considerable degree from the other bars in the same interval). The line traces a density that represents, approximately, the frequency of initial spin speeds falling within any not-too-small interval. From a frequentist point of view, it might be regarded as a genuine physical probability distribution.

A purist advocate of the dynamical approach to probability should, however, think of the density somewhat differently. Physical probability, they should maintain, exists only where there is a microconstant dynamics or something similar. The smooth distribution over actual initial conditions is a part of the foundation of such a probability, but it is not itself a self-standing physical probability distribution. Thus we should talk of a frequency distribution over the initial conditions, but not a physical probability or chance

20. This notion of actual-initial-condition smoothness, like the more general notion of density smoothness characterized in section 2, is relative to an outcome and a dynamics. There is no such thing as smoothness *simpliciter*, only smoothness with respect to such and such an outcome and such and such a dynamics.

distribution (unless the initial conditions are themselves produced by a microconstant process). In a slogan, the initial-condition distribution is to be given a *frequency-based*, but not a *frequentist*, construal.

I have focused so far on the orderly aspect of Bernoulli patterns—the stable long-run frequency. There is disorder, too, to explain. We can account for this disorder or statistical independence by invoking microconstancy together with the "weak independence" (my term) of the initial conditions, which is equivalent, for smoothly distributed initial conditions, to the smoothness of the joint distribution over those conditions. (The details are laid out in Strevens (2003), chap. 3.) Weak independence is, as the name suggests, a strictly weaker condition than independence, as it is compatible with correlations that full independence rules out. Microconstancy, then, generates from smoothness not only a striking form of order—stable frequencies—but also, within the framework of that order, a certain kind of maximal disorder.

To take the usual example, the disorder of *red* and *black* in a sequence of outcomes on a wheel of fortune can be explained by invoking the weak independence of the initial spin speeds. Such an explanatory strategy poses a problem, however, for a frequentist who confines their attention to the actual spin speeds producing the outcomes in question. What is wanted is the smoothness of the joint distribution over these speeds. For a series of (say) 200 spins, the joint distribution will have 200 dimensions, one for each spin. Only a single actual spin speed exists to provide a foundation for the shape of the distribution in each of these dimensions. There is obviously no prospect of interpreting smoothness by looking at the "statistics" of each of these isolated spins: just as two points cannot define a two-dimensional surface (in any physically meaningful way), so n points cannot define an n-dimensional surface.

What to do? One approach is piecemeal, explaining various aspects of Bernoulli disorder one at a time, using joint densities with many fewer dimensions than there are outcomes.[21] You might, for example, explain the lack of correlation between successive outcomes in a series of outcomes on the wheel of fortune as follows. Plot a two-dimensional bar graph, using the same subintervals as before, representing the frequencies of various pairs of spin speeds on successive trials. (A single bar might, for example, represent the frequency with which two successive spins have speeds falling into the intervals $[v - \epsilon, v + \epsilon)$ and $[w - \epsilon, w + \epsilon)$ respectively.) You may find some correlation in the speeds. Depending on the croupier, for example, you might find that a higher-than-average spin speed is more likely to be followed by another higher-than-average speed than by a lower-than-average speed. (Perhaps spin speeds run higher when the croupier has just had a coffee, or some particularly jaunty music is playing.) Because of the wheel's microconstant dynamics, however, this need not result in a correlation between the outcomes of successive spins: provided that the two-dimensional "frequency density" is roughly uniform over small regions, no such correlation will exist.[22] Smoothness of this density—or omitting the reference to a density altogether, roughly uniform distribution of actual pairs of successive spin speeds over small regions of the space of possible pairs—explains the complete lack of correlation between successive outcomes.

That is just one facet of the disorderliness characteristic of the Bernoulli patterns. Another is the lack of correlation between outcomes two places apart (the first and the third outcomes, the second and the fourth, etc.). The smoothness of a different two-dimensional frequency density can be recruited to explain this. And so on.

These are both 2-wise correlations. Clearly, there will be some point beyond which a lack of n-wise correlations cannot be explained by the smooth-

21. What follows is very much in the spirit of Reichenbach's characterization of disorderliness using different "selection functions" to capture facets of non-correlation one at a time (Reichenbach 1949, §30).

22. This is all a little roughly stated; I leave it to the interested (or concerned) reader to fill out the description, nailing down the meaning of terms such as "small".

ness of the corresponding n-dimensional densities, because there are too few actual spin speeds to determinately decide the issue of smoothness. Perhaps what we really want, when it comes to the explanation of disorder, we can get without going past this point.[23] But there is another approach to explaining disorder.

I have so far constructed my frequency-based distributions using only the values of initial conditions actually involved in causally producing the particular sequence of outcomes whose Bernoulli pattern is to be explained. It is quite in keeping with the frequentist philosophy, however, to cast the net more widely. We might look at the distribution of the actual initial speeds of all spins on the wheel of fortune in question, whether they belong to the sequence of interest or not. Or we might focus rather on the croupier, examining the distribution of the speeds of every spin they have ever made, on this wheel or any other of similar heft. In many cases these expanded reference classes will yield sufficiently many spins to decide the issue of smoothness for all the correlations (or rather, non-correlations) that we might seek to explain.

The obvious objection to broad-based frequency harvesting is that it goes far beyond the bounds of the causal process responsible for the patterns to be explained. How can an initial speed imparted to the wheel last week play a role in explaining the pattern of outcomes that the wheel produces today?

The answer is this: in explaining an outcome that is due to a certain configuration of initial conditions, it is acceptable to cite a much larger aggregate of which that configuration is a representative instance. We can explain the broken window, for example, by citing the riot, even though it was but a single brick tossed by a single rioter that caused the damage. The aggregate helps to explain because it is *entangled* with the actual cause, where "entanglement"

23. There will also be a point past which there will be too few *outcomes* to determine a fact of the matter as to whether there is a lack of m-wise correlation in the outcomes. The safe values of m, however, go rather higher than the safe values of n. There will be a zone, then, where we can discern a lack of correlation but where we cannot explain it by invoking smoothness in the actual initial conditions and microconstancy.

refers to a relation, comprising robust correlation and certain further properties, whose nature I have investigated in other work (Strevens 2008, 2012, 2014, forthcoming).

Entanglement is introduced in these discussions not only to account for the explanatory role of aggregates but also for a number of other purposes: to understand the nature of, and the proper explanation of, numerous high-level regularities; to account for exceptions to these regularities; to understand the causal relevance of mental content to behavior; and so on. A single relation does it all. But of course I do not expect you to be convinced by this capsule summary. I will, therefore, simply state my view: it is unobjectionable to explain the patterns in the outcomes of a particular series of spins on the wheel of fortune by pointing to the smoothness of the distribution of the initial conditions for a much larger set of spins to which the series in question belongs—provided, that is, that the initial conditions of this particular series are themselves smooth, or near enough as their statistics can determine.[24]

3.5 *The History of Initial Conditions*

Frequencies will get the job done, but there is a better way. That is to say, a frequency-based understanding of initial-condition smoothness is acceptable for explanatory purposes, but in the context of dynamical probability, it fails to exploit certain explanatory opportunities.

The dynamical approach looks for the explanation of a pattern of outcomes in the physics of the causal process by which they are produced. It is natural, then, for a dynamicist to ask concerning initial-condition smoothness: might the process that generates the initial conditions play a role in explaining the conditions' smoothness? Inspection of such processes would then lead to a deeper understanding of the Bernoulli and other probabilistic patterns.

24. The idea is developed for dynamical probability in particular in Strevens (2008), Part Four.

It might turn out that the causal process responsible for the initial conditions in question is itself irreducibly stochastic, and hence that the probability distribution over the initial conditions is one of irreducible probabilities—quantum probabilities, presumably (see note 1). But suppose not. Suppose, that is, that the initial conditions are themselves deterministically generated (or near enough) from further conditions, which causal precursors you might call the *ur-conditions*. Then we could perhaps explain smoothness by appealing to certain dynamical properties of the generating process, along with certain properties of the distribution of ur-conditions.

At this point you might wonder whether the croupier is not the only one spinning their wheels. What is the point of understanding the distribution of the initial conditions in terms of the distribution of ur-conditions, which will presumably in turn be understood in terms of ur-ur-conditions, and so on with the same interpretative problem arising at every iteration?[25] I hope I can convince you to suspend this entirely legitimate worry for a few pages while I look into the dynamical explanation of smoothness.

4. Explaining Smoothness Dynamically

The best approach to understanding smoothness, I suggest, is to focus on small perturbations or noise in the process by which initial conditions are produced. These minor but plentiful disturbances tend to flatten highly irregular distributions such as the paradigm of spikiness depicted in figure 4, eroding the "peaks" and filling in the "valleys". The result is a general propensity to smoothness.[26]

25. de Canson (forthcoming) lays out a version of this objection.

26. Poincaré (1914, §4.8) seems to have entertained a similar idea. His presentation of that idea is a little cryptic, and it appears to differ in important ways from mine. He writes, for example, that the smoothness of distributions tends to increase "over millions of centuries" (85), whereas the story I sketch below has smoothness created on the fly over and over, rather than depending on a glacially long-term progression. But there is at least some kinship between the two proposals.

In what follows, I will be satisfied with a bare sketch of this smoothing effect. (The sketch is based on an earlier treatment in Strevens (2013), §12.3, which provides some additional details.) My aim is simply to show you what a dynamical explanation of initial-condition smoothness might look like, and to identify some of the assumptions and limitations of the dynamical approach. The most salient of these is that—as anticipated above—the dynamical approach must make an assumption about the distribution of noise. To ground this posit, I will find myself turning back to frequencies.

Another limitation, but an entirely welcome one, is that the approach sets out to explain distributive smoothness only with respect to what I call the *standard variables*. The standard variables are those that we customarily use to quantify the physical character of the world: length measured in meters (or some other unit directly proportional to the meter), time measured in seconds, energy measured in joules, and so on. For every standard variable, there are infinitely many other ways of quantifying the same physical property (or as a metaphysician would say, determinable). We can measure distance in the standard way, or by using a function that takes the square root of the standard measurement—so that a distance that is four times further than another is represented by a number that is only twice as large.[27] These "gerrymandered" variables are non-standard.[28] If a standard variable is smoothly distributed, many highly gerrymandered non-standard variables quantifying the same determinable (though certainly not all such variables) will, of mathematical necessity, be non-smooth. It is impossible, then, to have a tendency to smoothness in all variables; we should not be looking for such a thing. Smoothness in the standard variables is quite enough to underpin the dynamical probabilities

27. Technically, a standard variable is a function from some real-world magnitude or determinable to the real numbers, or another appropriate mathematical space. It therefore has two aspects: a physical determinable, and a way of quantifying that determinable.

28. Unless they are employed as alternative quantifications of the same determinable, as sound pressure can be measured both in pascals or in decibels (the one being related logarithmically to the other).

we care about. So that is what I will endeavor to explain. (I will remark on the physical significance of the standard variables presently.)

On with the explanation, then. Consider the causal process generating the initial conditions of some microconstant process, for example, the process producing spins on a wheel of fortune. Let me call the initial conditions in question, such as the spin speeds, the "outcome conditions". The outcome conditions serve, then, as initial conditions of the microconstant mechanism of interest, but they are the endpoint, or product, of the process currently under examination. As foreshadowed above, I will call the input or initial conditions of this process the "ur-conditions". We are scrutinizing, then, the process that transforms ur-conditions into outcome conditions—for example, a physiological process that transforms biological and other states of affairs into particular spins on the wheel of fortune.

Suppose that, in the absence of what I am calling noise or perturbations, the process is deterministic, in the sense that any given ur-condition will produce a certain, fixed outcome condition. (Several ur-conditions may of course produce the same outcome condition.) The distribution of outcome conditions, then, is in isolation fully determined by the distribution of ur-conditions and the dynamics of the process. (Never mind for now what kind of distributions these may be.)

I have made no assumptions about either the shape of the ur-condition distribution or the nature of the dynamics. It might well be, then, that in isolation the distribution and the dynamics impose a distribution over the outcome conditions that manifestly lacks smoothness, such as the distribution shown in figure 4. My goal is to show that once we let in the noise, the resulting distribution will, under a wide range of conditions, be smoothed out.

Take any particular ur-condition. Without noise it generates, as I have said, a fixed outcome condition: the process is on rails. Randomly distributed perturbations will knock the process off its track, however; in a noisy environment, it will end up generating one of a range of outcome conditions. There will be a distribution over this range of possible outcomes, determined by the distribution of perturbations and the dynamics of the perturbing process. Call it the single-value distribution for that ur-condition.

Now suppose (I will return to this assumption shortly) that the single-value distribution for any ur-condition is smooth, that is, relatively flat over small intervals. ("Small" here is defined by the smoothness we are looking for in the outcome condition distribution, hence by the dynamics of the microconstant process lying downstream.) The outcome distribution as a whole will be made of the weighted sum of these many smooth single-value distributions, one for each ur-condition, with weights determined by the distribution over the ur-conditions.

It is easy to show that no matter what the weights, a weighted sum of smooth distributions is itself smooth.[29] Thus in the presence of noise, the distribution over outcome conditions is smooth.

Central to this proposed explanation of initial-condition smoothness is the smoothness of the single-value distributions. Under what conditions will such smoothness be the rule? A salient sufficient condition is, first, that the distribution over the perturbations is smooth—more precisely, that the distribution over the initial conditions of the perturbing process is smooth—and second, that the perturbing process itself has what might be called a "smooth and sensitive dynamics".[30] A dynamics is smooth if over small intervals it is approximately linear. It is sensitive if small differences in initial conditions result in greater differences among outcomes—or at least, differences that are no smaller.[31] Together, smoothness of the initial conditions and smooth-

29. A smooth distribution is approximately uniform over small intervals. Over any small interval, then, a distribution that is the weighted sum of many smooth distributions is the weighted sum of many approximately uniform distributions, hence is itself uniform. QED—bearing in mind that the deviation from uniformity of the weighted sum cannot be greater than the greatest deviation among the summands.

30. Also required is weak independence among the perturbations and between the perturbations and the ur-conditions, in order to derive weak independence of the outcome conditions.

31. For the purpose of evaluating sensitivity, by what metric should differences in initial

ness and sensitivity of the dynamics result in smoothness of the distribution over outcomes—thus, the smoothness of any given single-value outcome distribution.

The assumption of a smooth and sensitive dynamics is reasonable, I think, given our understanding of the physics of perturbation (or the physics of the many kinds of perturbation), at least when "small" really means small. It is not the case, of course, that a "rough" dynamics is impossible. But it is in a certain (loose!) sense unnatural: to effect non-linearity in the small normally requires a significant amount of causal channeling, that is, a causal structure that teases apart nearby initial states and treats them in physically heterogeneous ways. It is the sort of thing that we might find in computing equipment and other pieces of intricate nano-engineering, such as the evolved biological mechanisms for genetic transcription, but not in the relatively unsupervised goings-on that fall under the heading of noise.

Indeed, even in a microprocessor or a cell nucleus, the effect of perturbations on outcomes will much of the time be "smooth". Not all the time: these mechanisms are of course designed or evolved to resist in some respects the randomizing effects of noise. The smoothness of microdynamics is not ubiquitous. In situations where it is not to be expected, we ought not to expect it. But it is, nevertheless—so I claim—pervasive. Or at least, there is enough of it around to get us the initial-condition smoothness we are looking for.

I should perhaps add that naturally occurring chaotic processes are typically smooth. In a chaotic dynamics, small differences in initial conditions

---

conditions be compared to differences in outcomes? As a formal demonstration of the result under discussion would show, the answer is to be determined in conjunction with the meaning of all the other "size" terms in play. In particular, the degree of smoothness found in the distribution of perturbations and in the perturbing dynamics will determine how much sensitivity is required of the dynamics in order to produce a smooth single-value distribution over outcome conditions. It's all relative, then. An absolute measure of sensitivity, or for that matter of smoothness or of microconstancy (both of which depend on what counts as "small"), will have only heuristic or expository value, indicating the kind of property relevant to the smoothing effect.

result in big differences in outcomes, but by way of processes that tend to stretch out the space of the in-between conditions smoothly. Over small intervals, then, such a dynamics will tend to be roughly linear—though in the paradigmatic cases of chaos, highly sensitive, like the linear function that takes x to $100x$. (I will also say that, although chaotic processes are stirring exemplars of sensitivity, chaos in the technical sense is not necessary in order for noise to have its smoothing effect.)

Grant that we have established the prima facie credibility of the assumption of perturbing processes' dynamical smoothness and sensitivity. That leaves us with one further assumption to address, that of the smoothness of the distribution of the perturbers themselves, the noise. That topic will be taken up below.

Let me conclude the present section with a tally of the assumptions that go into the dynamical explanation of smoothness sketched here. First, noise is present everywhere; every causal process is thrown off course by small perturbations. Second, the dynamics of perturbation is on the whole smooth, that is, approximately linear over small intervals, and sensitive. Third, the perturbations that constitute noise are smoothly distributed and weakly independent of each other and of the ur-conditions (see note 30). From these three posits, and without supposing anything further about the distribution of the ur-conditions or the process by which they are transformed into outcome conditions, we can derive and thereby explain smoothness in the distribution of outcome conditions—that is, the smoothness of the initial condition distribution for the pertinent microconstant process.

All of this, I should add, ought to be understood in terms of the standard variables. It is relative to the standard variables that I claim that the assumptions hold true, and it is standard variables' tendency to distributive smoothness that is thereby explained. There are many non-standard variables relative to which the assumptions do not hold true and that show no tendency to smoothness. But there are also many non-standard variables—including "smooth" gerrymanderings of the standard variables—that do show such a tendency, in virtue of their satisfying the same assumptions. The smoothness explanation is not unique to the standard variables, then. And the explanation of their smoothness does not hinge on their standardness, which is to say on their perhaps parochial appeal to us, but on the mind-independent, physical fact that they do indeed satisfy the explanatory assumptions.

5. The Facts about Noise

If we take a dynamical approach to explaining the smooth distribution of noise, we will surely find ourselves assuming the smoothness of some other distribution, most likely over causally prior noise. We must cut through the circle. The unsubtle instrument for the job is frequency.

I propose, in this spirit, that we understand the smooth, weakly independent distribution over the perturbations that make up noise as a representation of the distribution of actual perturbations. In accordance with my remarks at the end of section 3.4, this distribution can encompass all the world's perturbations, not merely those actively involved in smoothing, without explanatory transgression.

Just as in the case of the distribution of the initial conditions of a microconstant process, the noise distribution need not be understood as a probability distribution in any sense that goes beyond the merely mathematical. In particular, it need not be interpreted as a representation of "objective probabilities" or "chances".[32] Its role is to help to explain the probabilistic patterns generated by microconstant processes, by smoothing the conditions that such processes feed on. As such, it is quite sufficient that it represent actual frequencies; it need not be dressed up in any further metaphysical finery.

The big picture I am painting is therefore as follows. We live in a world

32. The idea that such large-scale and fecund patterns of particular matters of fact might provide a "Humean" basis for a statistical law, without recourse to full-dress frequentism, is mooted by Loewer (2020) and developed in a somewhat more modest form by Hoefer (2019).

where actual noise is for the most part distributed smoothly and in a weakly independent way, meaning that its actual frequency distribution is smooth and individual perturbations are weakly independent of each other and of all other things (apart from their immediate effects, of course). Further, we live in a world of largely smooth and sensitive dynamics (relative to the standard variables). The dynamics by which perturbations affect the production of the initial conditions of microconstant processes are, in particular, typically smooth and sensitive. Noise, then, will have a characteristic effect on such initial conditions (expressed standardly): it will tend to iron out much or all of the roughness in their distribution. Consequently, we live in a world where the actual initial conditions of microconstant processes tend to be smoothly distributed (and weakly independent). This fact, together with microconstancy itself, explains the tendency for wheel spinning, coin tossing, die rolling, and other microconstant processes to produce Bernoulli-patterned outcomes.

Moving from explanation to metaphysics, I have suggested that where possible, we identify the probabilities in statistical models with the components of such explanations. The one-half probability for *red* on a wheel of fortune, then, will be constituted by the facts undergirding the microconstancy of the wheel's dynamics, and in particular the strike ratio of one-half for *red*, along with the tendency of the initial spin speeds on the wheel to arrive smoothly distributed—a tendency that is in turn constituted by (because it is explained by) the smoothness of the perturbing dynamics and a statistical fact, the smooth distribution of actual noise.

What does not play a role in either the explanatory or the metaphysical stories, observe, is any sort of distribution over the initial spin speeds' ur-conditions—the physiological and other conditions that generate spin speeds.[33]

33. The causal process by which the ur-conditions generate spin speeds is, by contrast, a part of the explanation of the Bernoulli patterns, because it is a part of what makes for the

Have we made progress? In order to explain a pervasive feature of our universe—that certain large classes of events exhibit the Bernoulli patterns—I have drawn on two kinds of facts: first, facts about physical dynamics, and second, facts about distributions of certain physical quantities, namely, the world's perturbations, or "noise".

From an explanatory perspective, this seems quite in order. Something has been explained in terms of something else that is causally prior. It is true that there are unexplained explainers. The smooth distribution of noise, in particular, is unexplained. But so it shall always be. Explanation must stop somewhere.

That said, it might not cease with a blanket assumption about noise's smoothness. There must surely be considerable scope for dynamically understanding ways in which that smoothness propagates itself—ways in which noise smooths the distribution of noise itself. Such explanations will never escape the need for an initial-condition distribution. Perhaps quantum mechanics will provide irreducible probabilities to do the job. But even if not—even if the ultimate initial-condition distribution must be given a frequency-based construal—we will have gained great understanding of the workings of the world.

From a metaphysical perspective, the ineliminable need for frequentist thinking might seem less satisfactory. We wanted to explain where physical probabilities, such as the one-half probability of *red*, come from. The answer has turned out to be: they come from other distributions. Or not quite—dynamics plays a crucial role, and indeed to my explanatory taste the more important role by far—but the distinctively *probabilistic* aspect of the physical probabilities, the character in virtue of which they conform to the probability

---

smoothness of the dynamics of the perturbing process. Consider, for example, a perturbation taking the form of a momentary impact that knocks the process off course before it has completed. The final result of the impact is determined jointly by the immediate effect of the impact and by the physics of the process that translates the state of the system at that point into an outcome condition.

calculus, comes from other distributions.

Might we turn back to typicality, or to Rosenthal's version of von Kries' range conception, thereby avoiding the appeal to a distribution over noise by declaring that, in place of physical features of the world, we can substitute mathematical facts (typicality) or metaphysical facts (the range theorist's insistence that dynamics alone is sufficient for the existence of probability)? Such maneuvers are perhaps viable, if they preserve the explanatory power of chance distributions. Whether that is true remains to be seen. In the interim, at least, I counsel a recourse to frequencies.

In that case, our physical probabilities will rest in part on other distributions. It might seem that we have failed to explain how physical probability makes its way into the world *ab initio*.

I resist this disappointing conclusion. There is an explanation. It has two distinct parts, frequency-based and dynamic. Probability comes to us, and charges our statistical models with something explanatorily vital, because the randomizing yet smoothing effect of the noise to be found in this world—in part a frequency-based matter of perhaps contingent fact—is transmogrified by dynamical properties such as microconstancy into something distinctively stochastic: probabilistically patterned outcomes.

So in dynamical probability, we find the dynamical facts that motivated Reichenbach's dissertation work integrated with the frequency-based facts that served as the foundation for his mature work—a fusion of the early and late Reichenbach that is markedly greater than its parts.

References

Abrams, M. (2012). Mechanistic probability. *Synthese* 187:343–375.

Albrecht, A. and D. Phillips. (2014). Origin of probabilities and their application to the multiverse. *Physical Review D* 90:123 514.

Allori, V. (ed.). (2020). *Statistical Mechanics and Scientific Explanation: Determinism, Indeterminism and Laws of Nature*. World Scientific, Singapore.

Butterfield, J. (2011). Less is different: Emergence and reduction reconciled. *Foundations of Physics* 41:1065–1135.

de Canson, C. (Forthcoming). Objectivity and the method of arbitrary functions. *British Journal for the Philosophy of Science*.

Engel, E. (1992). *A Road to Randomness in Physical Systems*, volume 71 of *Lecture Notes in Statistics*. Springer-Verlag, Heidelberg.

Filomeno, A. (2019). Stable regularities without governing laws? *Studies in History and Philosophy of Modern Physics* 66:186–197.

Frigg, R. (2011). Why typicality does not explain the approach to equilibrium. In M. Suárez (ed.), *Probabilities, Causes and Propensities in Physics*, pp. 77–93. Springer-Verlag, Heidelberg.

Hempel, C. G. (1965). Aspects of scientific explanation. In *Aspects of Scientific Explanation*, chap. 12, pp. 331–496. Free Press, New York.

Hoefer, C. (2019). *Chance in the World: A Humean Guide to Objective Chance*. Oxford University Press, Oxford.

Hopf, E. (1934). On causality, statistics and probability. *Journal of Mathematics and Physics* 13:51–102.

Keller, J. (1986). The probability of heads. *American Mathematical Monthly* 93:191–197.

von Kries, J. (1886). *Die Principien der Wahrscheinlichkeitsrechnung*. Mohr, Freiburg.

Lazarovici, D. and P. Reichert. (2015). Typicality, irreversibility and the status of macroscopic laws. *Erkenntnis* 80:689–716.

Loewer, B. (2020). The Mentaculus vision. In Allori (2020), pp. 3–29.

Maudlin, T. (2007). What could be objective about probabilities? *Studies in History and Philosophy of Modern Physics* 38:275–291.

———. (2020). The grammar of typicality. In Allori (2020), pp. 231–251.

Myrvold, W. (2012). Deterministic laws and epistemic chances. In Y. Ben-Menahem and M. Hemmo (eds.), *Probability in Physics*, pp. 73–85. Springer-Verlag, Heidelberg.

———. (2021). *Beyond Chance and Credence: A Theory of Hybrid Probabilities*. Oxford University Press, Oxford.

von Plato, J. (1983). The method of arbitrary functions. *British Journal for the Philosophy of Science* 34:37–47.

Poincaré, H. (1896). *Calcul des Probabilités*. First edition. Gauthier-Villars, Paris.

———. (1905). *Science and Hypothesis*. Translated by W. J. Greenstreet. Walter Scott.

———. (1914). *Science and Method*. Translated by F. Maitland. T. Nelson, London.

Reichenbach, H. (1949). *The Theory of Probability*. University of California Press, Berkeley, CA.

———. (2008). *The Concept of Probability in the Mathematical Representation of Reality*. Translated and edited by F. Eberhardt and C. Glymour. Open Court, Chicago. Reichenbach's doctoral dissertation, originally published in 1916.

Rosenthal, J. (2010). The natural-range conception of probability. In G. Ernst and A. Hüttemann (eds.), *Time, Chance, and Reduction: Philosophical Aspects of Statistical Mechanics*, pp. 71–91. Cambridge University Press, Cambridge.

———. (2016). Johannes von Kries's range conception, the method of arbitrary functions, and related modern approaches to probability. *Journal for General Philosophy of Science* 47:151–170.

Sklar, L. (1993). *Physics and Chance*. Cambridge University Press, Cambridge.

Strevens, M. (2003). *Bigger than Chaos: Understanding Complexity through Probability*. Harvard University Press, Cambridge, MA.

———. (2008). *Depth: An Account of Scientific Explanation*. Harvard University Press, Cambridge, MA.

———. (2011). Probability out of determinism. In C. Beisbart and S. Hartmann (eds.), *Probabilities in Physics*, pp. 339–364. Oxford University Press, Oxford.

———. (2012). The explanatory role of irreducible properties. *Noûs* 46:754–780.

———. (2013). *Tychomancy: Inferring Probability from Causal Structure*. Harvard University Press, Cambridge, MA.

———. (2014). High-level exceptions explained. *Erkenntnis* 79:1819–1832.

———. (Forthcoming). The explanatory role of aggregative properties. In C. K. Waters and J. Woodward (eds.), *Causation and Explanation in Biology*, volume 23 of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, Minneapolis.

Suárez, M. (2020). *Philosophy of Probability and Statistical Modelling*. Cambridge University Press, Cambridge.

Wilhelm, I. (In press). Typical: A theory of typicality and typicality explanation. *British Journal for the Philosophy of Science*.