

CHAPTER 9

.....

MECHANISTIC SOCIAL PROBABILITY: HOW INDIVIDUAL CHOICES AND VARYING CIRCUMSTANCES PRODUCE STABLE SOCIAL PATTERNS

.....

MARSHALL ABRAMS

THIS chapter explores a philosophical hypothesis about the nature of (some) probabilities encountered in social sciences. It should be of interest to those with philosophical concerns about the foundations of probability, and to social scientists and philosophers of science who are somewhat puzzled by the nature of probability in social domains. As will become clear below, the chapter is not intended as a contribution to an empirical methodology such as a particular way of applying statistics.

9.1. INTRODUCTION

.....

A relative of mine tells this story: While pulling on his overcoat after a day of interviews at an academic conference, a stranger walking past, noticing the name badge on his jacket, said “We’ve been looking for you!” The stranger and his colleagues

asked my relative to interview immediately, as they were to fly out in the morning. The relative eventually got the job and moved to a new city, remaining at the same university for over fifty years. Apparently, the subsequent course of his life depended in part on his pace and the speed with which he put on his coat in that moment. I'll give reasons think that cases like this, with sensitive dependence of individual outcomes on vagaries of individual circumstances, are the rule rather than the exception. I'll also argue that even though such idiosyncratic sequences of events appear peripheral to much work in the social sciences, their commonality is in fact what allows that work to proceed as it does. More specifically, sensitive dependence of *individual* outcomes on circumstances helps produce enough stability in *social* phenomena that they can be studied at all, allowing claims about probabilities of social phenomena to succeed. In the next sections I'll provide further details about the questions this chapter is intended to address.

9.1.1. Social probabilities

Social science research routinely makes claims about probabilities. These may be explicit claims that such and such outcome has a given probability, or may be represented by error terms or other terms, or may simply be implicit in assumptions. My interest here is in claims that one or another social outcome has (at least roughly) a certain probability. For example, consider the claim that among a group of ten- to twelve-year-olds living in a single-parent household in Seattle, there is a 25 percent probability of becoming a gang member by age 18 (cf. Hill et al. 1999). This is a claim about a relationship between two properties: living in a single-parent home during one period, and being a gang member during a later period. What is that relationship? It is certainly mathematical, but it's not merely mathematical, for social scientists use such probabilities to predict, explain, or otherwise make sense of social phenomena. Further, social scientists seem to believe that this sort of probabilistic relationship can sometimes be manipulated. For example, it's often thought that policy interventions and broad economic changes have the potential to alter such probabilistic relationships between antecedent conditions and outcomes; if successful, such manipulations can affect overall patterns of outcomes in a society. If social scientists did not think that probabilities of outcomes could be affected by policy, there could be no reasonable debate, for example, about whether the drop in crime in New York City during the 1990s was due to government crime-prevention policies or the healthy US economy in the during that period. Whether one thinks this question has been resolved or not, few would argue that such debates are senseless.

If the probabilities mentioned by social scientists are not merely mathematical, what is their nature? This a version of a longstanding question in the philosophy of science: What is an appropriate *interpretation of probability*—in this case, for certain probabilities in social sciences? Note that claims about probability in social sciences are often closely connected to claims about frequencies of outcomes, usually in

populations of people. Indeed, it's often claimed that the probabilities mentioned by social scientists just are frequencies. This is to adopt a *frequency* interpretation of probability for some contexts.¹ Although I'll give some of the (many) reasons that probabilities in social sciences should not be understood as frequencies, it's clear that an interpretation of probability for some central cases in social sciences should have a close connection to frequencies. In particular, I'll argue that social scientists frequently assume that frequencies of certain properties in populations are *stable*, which is to say that frequencies don't fluctuate wildly, but remain roughly the same over moderately short periods, and change only gradually over long periods. I think social scientists are often right to assume this, for the data often supports such assumptions. Why, though, should frequencies of outcomes for members of various social categories be stable? What explains that fact?

9.1.2. Individualism and Social Structure

Jencks et al. (1973) argue that factors they measured—family background, occupational status, educational background, and cognitive skill—played only a small role in determining income. Income inequality among randomly chosen men was just 12–15 percent greater than inequality between any of the categories studied. Jencks et al. suggest this result is not just due to the effects of unmeasured differences in ability:

Income also depends on luck: chance acquaintances who steer you to one line of work rather than another, the range of jobs that happen to be available in a particular community when you are job hunting, the amount of overtime work in your particular plant, whether bad weather destroys your strawberry crop, whether the new superhighway has an exit near your restaurant, and a hundred other unpredictable accidents. (Jencks et al. 1973, 227)

Consider, on the other hand, the view that certain factors themselves guarantee particular outcomes for people of particular types. Garfinkel (1981) describes a view inspired by Adam Smith, according to which those with given resources and ability who make certain choices will be rewarded by the free market in a way appropriate to those factors. The market guarantees this outcome. This view assumes that there is no luck so bad, or so good, that it can prevent a person from achieving his or her appropriate end.² Now, Garfinkel (1981) argues against the view that such individualistic properties are typically relevant to the kinds of outcomes studied by social scientists.³ Instead, structural factors such as the distribution of available jobs provide more appropriate explanations of social outcomes. Garfinkel nevertheless seems to suggest that given a set of structural factors, what happens to individuals sharing a set of relevant properties (e.g., educational background) is guaranteed.⁴ The view is not that other factors make no difference to a person's life, but that those within a relevant social category will, one way or another, end up with a particular outcome. Again, this view assumes that there is no luck so good or so bad that an individual will not experience the outcome determined for her

kind by prevailing structural conditions. It is implausible, however, that either a simple individual property or membership in a certain social category by itself guarantees an outcome. At the very least, some participants in state lotteries will find their income bracket changed, but the phenomenon is more general: As the example of my job-seeking relative and the discussion of Jencks et al. above illustrate, much of what happens in an individual life, including events with major consequences, can depend on idiosyncratic vagaries. I'll argue for this point further below.⁵

Now, it may be that for any given person in her particular, detailed circumstances, there is no other way that things could turn out than how they do. Nevertheless, in the aggregate, outcomes for people in different categories seem to occur with certain frequencies. Even if the world is fundamentally indeterministic, this indeterminism seems usually to average out to something like determinism in the gross physical and physiological processes that underlie social interactions. Why, then, do outcomes for members of different social categories often fall into certain frequencies in the aggregate? This question has no simple answer, but I will nevertheless offer what I think is a plausible answer.

9.1.3. Summary

We can summarize point of the preceding discussion as follows:

- Membership in few if any categories used in social sciences guarantees any interesting outcome.
- Much, perhaps almost all of what happens to each particular individual is plausibly determined by the details of his/her circumstances in a roughly deterministic manner.
- Much of what happens to individuals is such that minor variations in circumstances could have given rise to other outcomes.

I'll provide further support for these points below. If they are correct, how can it be that, for many categories used in social sciences, there are stable frequencies of outcomes? And what is the nature of probabilities in claims about the world in the social sciences?

The goal of this chapter is to suggest an answer to these questions for some central sorts of cases, by explaining how to apply a new interpretation of probability, far flung frequency (FFF) mechanistic probability (Abrams, forthcoming), to central cases in social sciences.⁶ (In what follows I'll often abbreviate "FFF mechanistic probability" by dropping "FFF" Context will make it clear when I intend "mechanistic probability" in a more general sense.) More specifically, my goals are to:

1. Further explain and motivate the need for an interpretation of probability appropriate for some social science contexts (§<1>9.2).
2. Summarize the core ideas of mechanistic probability, which depends in part on sensitive dependence of outcomes to initial conditions. (§<1>9.3).

3. Explain how FFF mechanistic probability could apply in some social science contexts, and argue that it's plausible that it does so, in part because individual outcomes do exhibit the requisite sort of sensitive dependence (§<1>9.4).
4. Connect my proposal to some related ideas, and describe areas for further research (§<1>9.5).

Note that (2) and (3) together constitute a fairly modest goal: To make plausible a hypothesis that has potential to resolve what seem to be difficult philosophical puzzles about probability and frequencies in social sciences. However, part of task (1) will involve arguing that no alternative hypothesis is even on the horizon.⁷ Given the lack of alternatives, it will be significant if the hypothesis of social mechanistic probability can be made reasonably plausible.

9.2. PUZZLES ABOUT PROBABILITY

9.2.1. Social Probability and Methodological Probability

I call probabilities such as those mentioned in the claim above about gang membership *social probabilities*, and distinguish them from another use of probability which I'll label by *methodological probability*; this distinction will be important here, even if it's not one that social scientists routinely make explicit. Social probabilities relate social conditions to social outcomes. They are part of what social scientists seek to discover, understand, provide evidence for, and so on. Methodological probabilities, on the other hand, arise from sources other than facts about social systems. For example, when a random-number generating algorithm is used to help sample members of a population, the probability of an individual being selected reflects, in part, the probability of the algorithm generating a number of a certain kind. The latter is a methodological probability. Similarly, when a Monte Carlo simulation is used to learn about a social system, the random numbers used to determine parameters on different simulation runs reflect methodological probabilities set by the researcher.⁸ Note that sampling and Monte Carlo methods are used to make inferences about the world, and conclusions which these method help to justify may often concern social probabilities.⁹

Probabilities in error terms can reflect both methodological and social probabilities. For example, one kind of error is due to the fact that a sample might not accurately represent a population. Some of this error has to do with the fact that the random numbers used to generate the sample need not be distributed precisely according to an algorithm's underlying probability distribution (just as flipping a fair coin ten times need not produce five heads and five tails). However, the error might also reflect idiosyncrasies in the sample, as when it happens that there are

correlations between relevant properties of sampled individuals, which differ from correlations between the same properties in the population. The correlation in the sample then reflects probabilities in the population, as well as those generating the random choices of individuals. Although the distinction between social and methodological probabilities raises subtle issues deserving of further discussion, I don't need to address these issues here, as long as it is clear that some claims about social phenomena concern probabilities that are purely social. For example, my earlier claim about the probability of gang membership has nothing to do with random-number generators or any other nonsocial source of probabilities, even if the evidence for it might depend partly on methodological probabilities. In the rest of the chapter my focus will be on social probabilities.

9.2.2. The Need for an Interpretation of Probability

First note that from the point of view of mathematics, a probability is any function $P()$ which satisfies these axioms concerning subsets of a set Ω :¹⁰

1. Probabilities range from 0 to 1: $0 \leq P(A) \leq 1$, for any subset A of Ω .
2. An outcome which is certain to occur has probability 1: $P(\Omega) = 1$.
3. The probability of either A or B occurring, where A and B are mutually exclusive, is the sum of A 's and B 's probabilities: $P(A \cup B) = P(A) + P(B)$ for any two subsets A and B which have no elements in common.

A function satisfying these requirements is a *probability measure*, and the number it assigns to a set is the set's *measure*.¹¹ (This use of "measure" carries no implication that empirical measurement is involved.)

Now consider the question "What is the probability of the portion of the page above this printed sentence?" This is a perfectly reasonable question from the point of view of mathematics: If we divide the page up into a finite number of basic subsets, proportion of area of the page satisfies the preceding axioms, and thus counts as probability in the mathematical sense. The question will no doubt sound odd to most readers, however, because the sense of probability used in science (and everyday life) typically goes beyond what mathematics requires. Exactly how it does so depends, somewhat, on context. At the very least, probability in this sense:

1. Concerns situations, events, states, and so on, that might occur in the world.
2. Satisfies mathematical axioms of probability.
3. Satisfies other criteria that capture roles of probability in scientific contexts.

There is disagreement about what this last item should include. However, probability plays different roles in different contexts, so it seems implausible that a single set of criteria will be universally appropriate. At the very least, we can distinguish probabilities that in some sense capture facts about degrees of confidence, and probabilities that capture facts about relationships in the world. The former are called *subjective* or *epistemic* probabilities, the latter, *objective probabilities*. I'll argue

that an objective interpretation of probability is needed to make sense of claims about social probabilities. In addition to satisfying probability axioms, I claim that the following somewhat loose criteria are important for many (objective) probabilities in social sciences.¹²

1. Ascertainability: It should be possible, in principle, to determine the numerical values of probabilities.
2. Objectivity: Probabilities should be constituted only by facts about states of the world; epistemic factors such as belief or justification are not involved.
3. Explanation of frequencies: That an outcome *A* has the probability it does should help explain why frequencies of *A* are what they are.
4. Distinguishability of nomic regularities: Closely related to the preceding is that an interpretation of probability should be able to make at least a rough distinction between instances of nomic regularities, that is, cases in which frequencies are what they are for some systematic reason, and instances of accidental regularities, in which frequencies just happen to turn out the way they do for no particular reason (Strevens, 2011).

The importance of ascertainability should be obvious. A simple example will clarify the general ideas behind the other three criteria.

Consider these two procedures for generating a set of heads and tails from a coin dropped on a table 1,000 times.

1. I toss an evenly-weighted coin 1,000 times in the usual way; it comes up heads on 493 tosses.
2. I decide to stop strangers on the street, and get three people to give me a number between 0 and 9. This will produce a 3-digit number. I'll then drop a coin from a small height onto a table 1000 times, doing so carefully in order make the number of heads equal whatever 3-digit number resulted from my street interviews. I follow the plan: The strangers give me the numbers 4, 9, and 3, in that order, so I drop the coin 1000 times from a centimeter above the table, producing exactly 493 heads.

It's clear that the explanation of the frequency of .493 must be different in the two scenarios. The frequency of .493 fits a systematic pattern in the first instance in a way that it doesn't in the second: There is something about the coin and the standard coin-tossing strategy that explains the fact that the frequency of heads is near .5 in the first scenario. In the second scenario, it's not clear that there is any systematic explanation of the frequency of .493. At the very least, there doesn't seem to be any particular reason that the frequency is near .5. Note that in either case, what explains the frequency concerns objective factors in the world, rather than facts about my subjective or epistemic state.¹³

Here I'll briefly mention some well-known interpretations of probability. My goal is not to do justice to them, but merely to use some of their most obvious drawbacks as a way to clarify the motivation for mechanistic probability.¹⁴ Note that my discussion in this section is largely independent of debates about Bayesian versus

frequentist statistics. Though these two schools of statistical methodology do draw their initial motivation from Bayesian and frequency interpretations of probability, in practice their use is not firmly tied to either interpretation.

First consider a simple *finite frequency* interpretation of probability, according to which probabilities in science are identified with frequencies of outcomes—for example, gang membership—in an actual set¹⁵ of occurrences—for example, the set of actual teenagers in Seattle during the 1990s. However, this interpretation implies that *improbable* combinations of outcomes cannot come to have a high frequency by chance—if they did, they would by definition not be improbable (Hájek, 1996). That is, finite frequency doesn't distinguish nomic from accidental regularities, since whatever outcomes occur, for whatever reason, are treated identically. For example, there is no difference between the two coin tossing scenarios above from the point of view of a simple finite frequency theory.

Long-run frequency interpretations define probability as frequency in large, counterfactual sets of events, or as the limit of frequencies in a counterfactual infinite sequence. However, in addition to deep problems that I will pass over (Hájek, 1996, 2009a), neither long-run frequency nor finite frequency interpretations can explain frequencies in actual events, since the sets of events whose frequencies these interpretations trade on merely contain the actual events to which probabilities are attributed. In the case of a simple finite frequency theory, a probability simply is the frequency in an actual set of events; it therefore can't explain that same frequency in that same set of events. In the case of a long-run frequency theory, a set of actual events is supposed to be a subset of the full counterfactual sequence of events that define a probability. Putting aside problems about the explanatory power of such counterfactual frequencies, a frequency in a set or sequence of events will not explain a frequency in a proper subset of those events, without additional substantive assumptions—for example, assumptions about processes which sample from the larger set/sequence. Such assumptions would be difficult to justify in the general case, and are not usually part of a long-run frequency interpretation of probability.¹⁶

Best System Analysis *chance* interpretations of probability (Lewis, 1980, 1994; Loewer, 2001, 2004; Hoefer, 2007) have some popularity among philosophers currently because they do a good job of addressing certain philosophical problems. These theories are too elaborate to summarize quickly. One core idea is, very roughly, that an objective chance exists whenever assuming its existence plays a simplifying role in an ideal, ultimate set of scientific theories—which are those that succeed well in summarizing all actual events in the universe, past, present, and future. What Best System theories don't do well is allow probabilities to explain frequencies, for Best System chance probabilities are defined in terms of whatever states the actual world has. We then again end up explaining frequencies in terms of themselves.¹⁷

Bayesian interpretations define probabilities as degrees of belief or epistemic relations. Apart from the fact that these interpretations don't satisfy *objectivity*, Bayesian probabilities can only reflect and track frequencies in the world; they don't explain them. A researcher's degree of confidence in an outcome for Seattle children will not explain that outcome, in general.

Propensities are postulated indeterministic dispositions, modeled on deterministic dispositions.¹⁸ According to some common philosophical theories concerning causal properties, a lump of salt sitting in a salt shaker has deterministic *disposition* to dissolve when placed in water at room temperature, even if it is never placed in water. Propensity is a proposed extension of this notion of disposition. For example, some authors claim that an evenly weighted coin has an indeterministic disposition of strength 1/2 to produce the outcome heads when tossed in the usual way. Note that since propensities are supposed to be dispositions, and dispositions are supposed to involve causal properties, propensities would have potential to provide causal explanations of outcomes and frequencies. However, because there's not much more to propensity theories than what I have just said, many philosophers argue that propensity theories are too vague and unmotivated to have explanatory force. There are other objections to propensity theories as well (Eagle 2004). My view is that at the very least, attempting to give propensities a legitimate role outside of fundamental physical contexts requires an incoherent conception of causation, tantamount to allowing a single token event to simultaneously cause and fail to cause an effect (see Abrams, 2007, forthcoming).

9.2.3. Why are Social Frequencies Stable?

Social scientists often assume that although social conditions change, frequencies don't usually fluctuate wildly. Sometimes there is evidence for such stability. For example, although per capita welfare caseloads in the United States between 1970 and 2005 have fluctuated from year to year, with a precipitous drop between 1995 and 2002, a graph of caseloads from year to year appears roughly continuous (figure 9.1). There are no years in which the caseload shows a large drop or jump, and where there are large changes, they are usually followed or preceded by changes in the same direction. Even without data to prove that patterns at the level of populations don't fluctuate wildly, social scientists regularly assume such stability, even if tacitly. For example, Hill et al. (1999) found correlational patterns between characteristics or circumstances of juveniles aged ten to twelve, and prevalence of gang membership in the same cohort a few years later. The authors suggest possible interventions to reduce gang membership. It would make no sense, though, to consider interventions on a system unless data collected earlier were likely to be somewhat representative of patterns at later periods of time. For that matter, it seems unlikely that most social science research on groups of individuals would have any interest at all, if we didn't think that conclusions were in most cases roughly projectable into the past and the future. Further, when there are radical changes in frequencies from one time period to another, the response thought appropriate is usually to look for an explanation of the change, whether in social conditions, the physical environment (e.g., an earthquake), or in research methodology. The default assumption of relative stability is simply a precondition for much work in social sciences. It is an assumption, moreover, which plausibly has borne fruit; to the extent that social science research has been successful, the stability assumption is empirically justified.

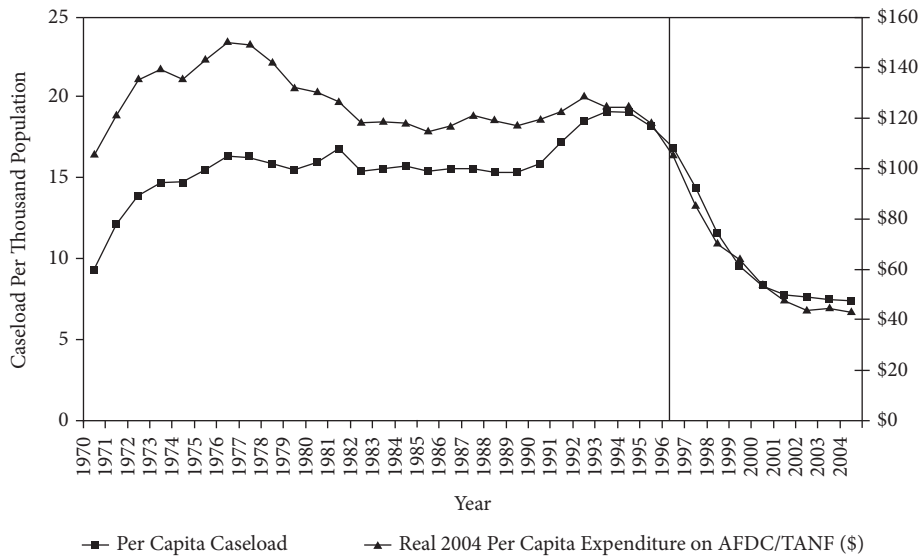


Figure 9.1 Per capita AFDC-TANF caseload and expenditure 1970–2004. Sources: U.S. Department of Health and Human Services, “Indicators of Welfare Dependence,” *Annual Report to Congress*, Appendix A, Table TANF4, Washington, D.C.: Office of the Assistant Secretary for Planning and Evaluation, 2006; and U.S. Department of Commerce, *Statistical Abstract of the United States*, U.S. Census Bureau, Washington, DC, 2006, Table [Population](#). (From Moffitt 2008).

That frequencies are stable is such a pervasive background assumption that it may seem that no explanation of stability is needed. But patterns in populations such as the one that Hill et al. studied are the result of numerous decisions by people with different characters, in circumstances that differ both in large and small ways. Notwithstanding common intuitions, there’s no obvious, good reason that this should result in stable frequencies rather than wild, chaotic fluctuations. When pushed, I think the first explanation of stability that comes to mind, for many, is something like this:

(1) Some children and other people act in one manner, others act in another. At some level of description, there are no particular patterns to what people do or don’t do. As long as you aggregate a sufficient number of individuals, the average behaviors will remain stable over time, producing stable frequencies concerning gang membership.

This argument is incomplete at best. Why do the underlying behavioral causes of measured outcomes comport themselves so as to maintain stable frequencies? That there is no particular pattern does not imply that the underlying behavioral causes should manage to sort out, year after year, so as to keep frequencies of outcomes from changing radically. It would be different if there were a reason to think that there were stable patterns in the underlying behaviors producing the observable frequencies. For example, if there were objective probabilities that the underlying behaviors tended to follow for some principled reason, and the distribution was such that the effects of these behaviors would be stable, such probabilities could

explain stable frequencies.¹⁹ But the mere assumption that there are no particular patterns to the underlying behaviors doesn't justify the claim that there are probabilities governing those behaviors. "No particular pattern" is consistent with frequencies following an unsystematic chaotic pattern, or being stable, or anything in between.

This suggests another possible answer to the question of why social frequencies are often stable:

(2) Twelve-year-olds are like quantum phenomena: Each child in the same economic class is fundamentally indeterministic, and has exactly the same tendency to join a gang as any other member of the class—independent of any of his/her characteristics or any circumstances encountered. (If quantum mechanical probabilities are propensities, then the claim is that there is a determinate propensity for gang membership given being in a certain economic class.) That would be why the frequency of gang membership within each class is stable over time. On this view, once we determine the child's economic class, there is nothing else to be learned which might be relevant to gang membership.

If this were true, economic class would be the *only* factor determining whether a child joined a gang or not. There would no point in investigating other factors such as race, number of parents, availability of drugs, quality of teachers, and so on. Nor could it be relevant to investigate the effects of early childhood trauma, fetal environment, brain chemistry, or genetic influences. Individual circumstances would be irrelevant to individual outcomes as well. It could not make a difference, for example, whether an existing gang member moved in next door, or whether a child was traumatized when an inadequately maintained bridge collapsed. Thus the present view, according to which a particular social category gives rise to probabilities of outcomes all by itself, precludes any further scientific research of any kind, social or otherwise.

Consider, finally, this explanation of stable frequencies:

(3) The relationship between properties of children and later gang membership is deterministic, or effectively deterministic (i.e., fundamental indeterminism has almost no noticeable effect on outcomes). In principle, if we could learn enough about a given child, her genes, and the history of her interactions with the world, gang membership would be determined, or would at least have an irreducible probability very near 0 or 1. However, at some point in the past, initial conditions were such as to produce frequencies like those we see to day, and the entire social system is such as to roughly reproduce frequencies from year to year. That is, society has a causal structure that makes it carry forward frequencies from the recent past (cf. Millikan, 2000). Stable social frequencies are thus analogous to a distribution of rock sizes over a hillside during a landslide: The distribution of rock sizes across the hill at one level is roughly the same as that at a slightly higher level a moment earlier because the interaction with gravity and the underlying shape of the hillside preserves, roughly, the ordering of rocks across the hillside. (This might not be true for balls dumped into the top of a pinball machine.)

I don't have a general theory about what kind of causal structures might have this property of carrying forward frequencies, and even if we did have such a theory for social systems, a mystery would remain: Why should the determinants of so many properties of so many social systems ~~should~~ have this kind of special, frequency-preserving causal structure? I'll argue instead that social systems in effect repeatedly *recreate* similar frequencies (except for important special cases, section 9.5.2).

Although I will assume below that social systems are effectively deterministic, I'll argue that for the most part social frequencies are stable because social systems have a special causal structure, "bubbliness," which I'll describe below.²⁰ This "bubbly" causal structure helps generate frequencies in a way that's largely insensitive to variations in recent conditions. That is, rather than stable frequencies being the result of a system that reproduces frequencies from the recent past, I'll argue that many social systems have a causal structure which makes it difficult for any but a narrow range of frequencies to be produced over a wide range of initial conditions. It is this causal structure that is at the core of the notion of FFF mechanistic probability (Abrams, forthcoming), and I'll argue that FFF mechanistic probability is an appropriate interpretation of probability for many contexts in the social sciences. As some propensity interpretations are supposed to do, FFF mechanistic probability is defined in terms objective aspects of the world and helps explain stable frequencies. FFF mechanistic probability is one of three closely related objective interpretations of probability that have recently been proposed, as mentioned above, including also Strevens's "microconstant probability" (Strevens, 2011) and Rosenthal's "natural range" conception of probability (Rosenthal, 2010). I call all of these interpretations "mechanistic probability" interpretations, because they depend on the same causal structure, but there is no established term for such interpretations at present.²¹

9.3. MECHANISTIC PROBABILITY

In this section I'll sketch the core concepts of mechanistic probability using the idea of a wheel of fortune. In the next major section I'll explain how to apply mechanistic probability in common social science contexts.

9.3.1. Why Wheels of Fortune Generate Stable Frequencies

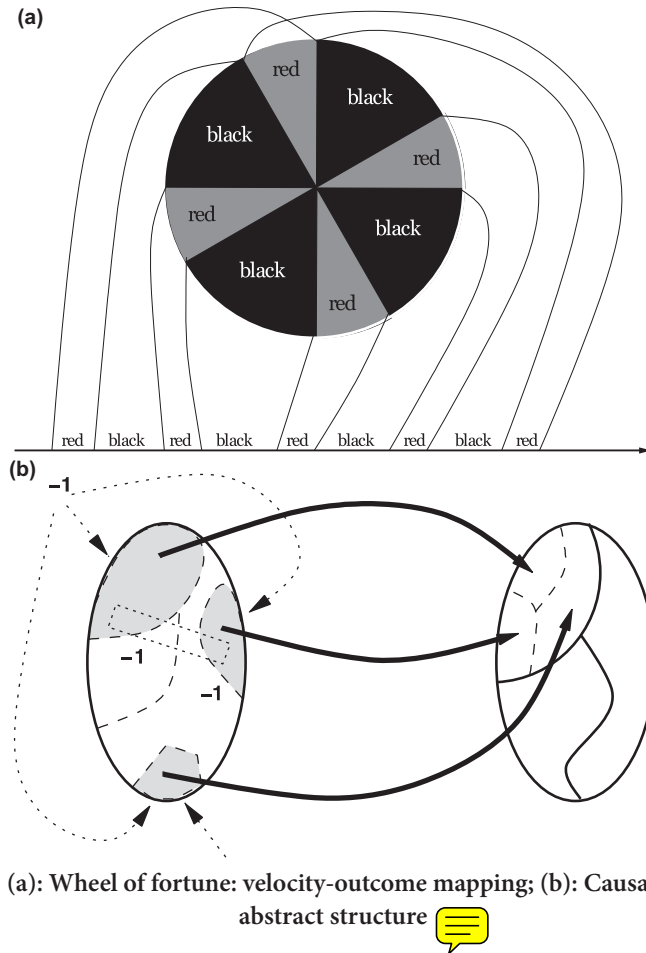
Consider a wheel of fortune, a simplified roulette wheel in which the outcome is indicated by the wedge nearest to a fixed pointer when the wheel stops. No ball is involved. If such a wheel is repeatedly given a fairly fast spin, the frequency of an outcome in a set of such spins will usually be close to the outcome's proportion of the circumference of the wheel. This is so in general, no matter who the croupier is. Casinos depend for their livelihood on such facts. It's natural to think there is in

some sense an objective probability of an outcome equal to its proportion of the circumference—even in unusual cases in which a run of good or bad “luck” produces a frequency that departs significantly from that probability. I want to explain why the wheel of fortune exhibits stable frequencies, and lay the groundwork for defining an interpretation of probability that will apply to the wheel of fortune and other systems.

The wheel of fortune is what I call a *causal map device*, an effectively deterministic device which, given an initial condition, produces an outcome. A croupier gives the wheel a spin—imparts an angular velocity to it—and Newtonian physical processes take over. The wheel then produces a particular outcome, red or black, depending on which wedge is nearest to the pointer when the wheel stops. Figures 9.2(a) and (b) will help clarify what it means to say that the wheel of fortune is a causal map device.

Note that we can divide the set of initial conditions, consisting of angular velocities possible for human croupiers, into those regions which lead to the red outcome, and those which lead to the black outcome; Figure 9.2(a) provides a schematic representation of this point. Velocities are represented on the horizontal axis, increasing as we move from left to right. The velocities are divided into regions of velocities leading to red or to black (marked by “red” and “black” above the line); lines rise from the borders between regions, or intervals. For example, the leftmost interval marked on the horizontal axis represents a set of similar velocities, from lowest to highest. These are velocities that would cause the wheel to stop with the pointer indicating the red wedge near the top of the wheel (between 11:00 and 12:00). The fact that the curved lines lead to the uppermost red wedge is intended to convey this relationship. Similarly, the endpoints of the next velocity interval to the right are starting points for curved lines that rise to the black wedge that’s adjacent and the left of the red wedge just mentioned (i.e., the black wedge between 9:00 and 11:00). The idea is that initial velocities in this black region would cause the wheel to spin a little bit longer, so that the pointer will end up pointing at the next wedge after the uppermost red wedge, that is, at the black wedge next to it. This pattern continues as velocities increase along the horizontal axis. Velocities in the rightmost interval marked on the diagram cause the wheel to spin a full revolution further than do velocities in the leftmost interval, so that these greater velocities also produce a red outcome by causing the pointer to indicate the red wedge between 11:00 and 12:00. Thus the curved lines from both the leftmost and rightmost velocity intervals lead to the edges of this same wedge.

Figure 9.2(b) is a general representation of the abstract structure of a causal map device. The oval on the right represents a space of possible output effects from a causal map device; these outputs are divided into three outcomes, *A*, *B*, and *C*. (For the wheel of fortune, particular locations on the wheel when it stops count as outputs, which are divided into only two outcomes, red and black.) On the left side of the diagram, a space of initial conditions is divided up into three sets A^{-1} , B^{-1} , and C^{-1} , which are the sets of initial conditions which would cause outputs in *A*, *B*, and *C*, respectively.²² Recall that in the case of the wheel of fortune, different velocities



Figures 9.2 (a): Wheel of fortune: velocity-outcome mapping; (b): Causal map device: abstract structure

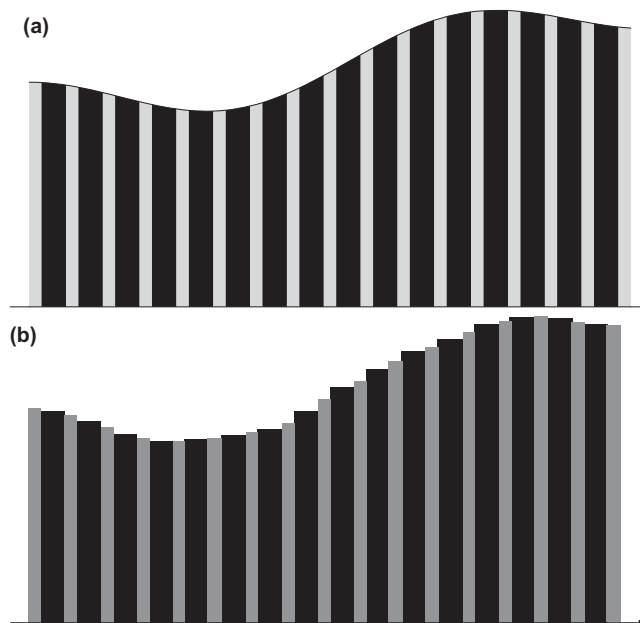
all produced the red outcome. Further, velocities in disjoint subsets—different intervals—caused outputs corresponding to points in distinct subsets of those outputs that counted as red outcomes; there the subsets were distinct red wedges. In the same way, figure 9.2(b) shows that initial conditions in different parts of A^{-1} (the three gray areas) produce outputs in different parts of the A outcome region. This is indicated by dark arrows leading from the three parts of A^{-1} to subsets of A divided by dashed lines.

The wheel of fortune is a “bubbly” causal map device: Its input space can be divided into many “bubbles”, small regions containing initial conditions leading to all outcomes—in this case, red and black.²³ More specifically, consider any adjacent pair of intervals, one containing velocities leading only to red, and one containing velocities leading only to black. The union of those two intervals counts as a bubble, because it contains initial conditions, or inputs, leading to each outcome.²⁴ For example, the two leftmost intervals along the horizontal axis in figure 9.2(a) constitute a bubble. Intuitively, in a normal wheel of fortune, such a bubble is small, in that the distance between its smallest and largest velocity is a small fraction of the distance

from one end of the input space to the other. (This remark is only intended to be suggestive, though. I'll give a fuller characterization of bubble size shortly. At this point an informal presentation with a few slightly inaccurate remarks will make it easier to follow the more systematic description of mechanistic probability below.)

For a given croupier, we can graph the number of spins she imparts at any given velocity. Figure 9.3(a) illustrates this idea for a croupier who spins the wheel many, many times, but who tends to give the wheel faster spins more often than slower ones. The height of the curve at each point reflects the number of spins the croupier gives at the velocity represented by that point on the horizontal axis. This curve is called a density curve, since it represents the “density” of spins at each velocity. (On the other hand, the graph in figure 9.3(a) represents the croupier as generating spins at *every* velocity possible for human croupiers; this is of course unrealistic. I'll come back to this shortly.)

To summarize what's to come: Very roughly, it can be proven that if bubbles are small and if the slope of a croupier's density curve over inputs is not extremely steep in any region, the frequency of red will be close to the proportion of the input space occupied by points leading to red (Strevens, 2003; Abrams, forthcoming). The same claim holds for the black outcome. In other words—again very roughly—if the red and black intervals can be paired so that each pair (bubble) is small compared to the entire range of possible spin velocities, and if the height of the croupier's density curve is roughly the same in both parts of each bubble, then the relative frequencies of outcomes will be close to the relative sizes of red and black wedges. The relevant proofs provide at least a partial explanation of what it is about the wheel of fortune



Figures 9.3 (a): Hypothetical velocity distribution for a croupier (continuous); (b): Hypothetical velocity distribution for a croupier (discontinuous)

that gives rise to the phenomenon of stable frequencies matching proportions of the input velocity space.

An understanding of the general idea of the proofs can be gotten from figure 9.3(a). Notice that since the height of the curve at a particular velocity represents the number of spins at that velocity, we can think of the area of the region under the curve and above a particular interval as representing the total number of spins in that region of velocities, with larger areas containing more spins. Roughly, we multiply the width of a region by the average height of the curve over it to get the approximate number of spins at velocities in the region. (I'll give a more careful account of this idea shortly.) Since in each pair of contiguous red and black regions, heights don't differ much, the ratio between the number of spins in the two corresponding velocity ranges is then close to the ratio between the widths of the two ranges. As this is true for each such pair—each bubble—the ratio between number of red and black outcomes will also be close to the ratio between the summed velocity intervals leading to red and the summed intervals leading to black. In this case, that ratio is close to the ratios between sizes of red and black wedges. It can be argued that this sort of pattern is a large part of what explains the stable frequencies found in many mechanical casino games.²⁵

9.3.2. Mechanistic Probability: Preliminaries

Of course, no real croupier will ever spin a wheel of fortune at every point in a continuous interval of velocities. It makes more sense to replace figure 9.3(a) with a step-shaped curve, as in figure 9.3(b). Here the area of each rectangular region represents the number of spins in the velocity interval, which is represented as the lower edge of the rectangle. A more abstract version of this idea will be part of the characterization of FFF mechanistic probability, and will help illuminate the reasons that wheels of fortune—and social systems, I'll argue—exhibit stable outcomes frequencies. Let me first provide some background.

Recall that a mathematical probability measure is nothing more than a function which assigns numbers between 0 and 1 (inclusive) to sets in such a way that axioms like those given above are satisfied. This idea plays an essential role in the account of mechanistic probability, so a concrete illustration may be useful: One possible probability measure on the input space of the wheel of fortune uses the fact that velocities are represented by real numbers. Let's call the minimum and maximum velocities possible for a human being v_0 and v_1 . Suppose that our probability function P assigns 1 to the set containing all of the velocities possible for a human croupier ($P([v_0, v_1]) = 1$). More generally, consider any interval containing all velocities lying between two real numbers, v_i and v_j , which lie between the minimum v_0 and maximum v_1 , where v_j is larger than v_i . One way to assign a probability to an interval of velocities is to take its probability be the ratio between the interval's length and the length of the set of humanly possible velocities. In other words, let the probability of any interval within the range bounded by the minimum velocity v_0 and the maximum v_1 , be the difference between the interval's largest and smallest velocities,

divided by the difference between the largest and smallest velocities which are humanly possible:

$$P([v_i, v_j]) = \frac{v_j - v_i}{v_1 - v_0}$$

This is the core idea of what's called a "normalized Lebesgue measure" for the wheel's input space.²⁶ It corresponds to the intuitively natural way that we measured "widths" of velocity intervals above.

Although normalized Lebesgue measure is in some sense a natural way of thinking about probabilities for sets of initial velocities, I intend it mainly to illustrate the application of a mathematical probability measure to an input space. Many such mathematical probability measures could be assigned to the input space of a wheel of fortune. We might, for example, give higher probabilities to regions containing large velocities, or regions that happen to lead to black outcomes. These are only some of the simplest examples. Whether any of these mathematical probability measures correspond to a probability measure in an objective sense—whether a particular mathematical function captures mathematical aspects of an objective probability function that satisfies the desiderata above—is a further question. Note, however, if a given mathematical probability measure *did* correspond to probabilities of inputs to the wheel of fortune in some objective sense, it might be appropriate to define the probability of an *outcome* as the probability of those inputs which can produce it. Thus, suppose, for example, that a normalized Lebesgue measure on the wheel of fortune's input space captured mathematical aspects of an objective probability measure concerning sets of possible initial conditions. Then there would be an objective probability of red equal to the total probability of the velocity intervals whose points could produce the red outcome.²⁷ In figures 9.3(a) and (b) this probability would be proportional to the summed widths of the velocity intervals with red (i.e., gray) rectangles over them.

For now, let us assume—pretend, if you like—that we are given a mathematical probability measure P on the input space of initial conditions, for a causal map device such as a wheel of fortune. This assumption will help us get clear on ideas, though in the end I'll dispense with it, and it doesn't matter at this point which function the probability measure is. It might be normalized Lebesgue measure, or it might be some other mathematical probability measure. Note that whether or not there is an objective correlate to a given mathematical input probability measure, we can nevertheless define a derived mathematical probability measure on outcomes as we did in the preceding paragraph: That is, we can define the mathematical probability of an outcome as the probability of those inputs which can lead to it. The fact that this relationship mirrors a causal relationship between inputs and outcomes will be significant below, but at this stage the point is just that it provides a way of defining a *mathematically* legitimate probability measure over outcomes. (Shortly, I'll explain how we can *choose* an appropriate input probability measure so that outcome probabilities will usually correspond to stable frequencies.)

Now, assuming a given probability measure on the input space, the microconstancy inequality theorem (see appendix to this chapter) says that for a given distribution of initial conditions, the difference between the relative frequency of outcome A and the input measure of initial conditions which lead to A is constrained by *bubble-deviations* and bubble measures. I'll explain what a bubble-deviation is in a moment, but first let me give the context in which they appear: What the theorem says, roughly, is that when bubble-deviations and bubble measures are small, outcome frequencies will usually be close to outcome probabilities which have been defined by an input measure in the way just considered. (That is, an outcome probability is the probability of inputs that could cause it.) This will mean that any input measure that can make bubble-deviations and bubble measures small for most real distributions of inputs, will also make outcome frequencies close to outcome probabilities defined in terms of that input measure. I'll present these points in more detail below.

The notion of a bubble-deviation is a little bit abstract, but it can be viewed as an analogue of slope for a distribution of initial conditions. Bubble-deviation is defined in terms of an input measure of a bubble and a distribution of inputs—for example, of spins of a wheel at the possible velocities. Unlike slope, bubble-deviation doesn't depend on using a real-numbered scale. (Velocities fall on a real-numbered scale, and we can use that fact to define probability measures such as Lebesgue measure, but bubble-deviation doesn't depend on this fact.)

To understand the idea of a bubble-deviation, look at figure 9.4(a), which here represents part of a distribution over initial spin velocities for one particular croupier. More specifically, 9.4(a) displays, in schematic form, one pair of black/red bars from figure 9.3(b). The bottom edges of the boxes represent two velocity intervals, making up one bubble. The bottom edge of the right-hand rectangle then represents a set a in the input space—a set of inputs that all would lead to one outcome, red. Thus a could be, for example, the rightmost interval along the horizontal axis in figure 9.2(a). The bottom edge of the left-hand rectangle (here \bar{a}) would then correspond to the next interval to the left in 9.2(a).

In figure 9.4(a), width represents input measure, and the right-hand rectangle represents the croupier's total number of inputs in a . The rectangle's height represents the average number E_a of inputs in a , where this average is computed using the input measure p_a ; in other words, $E_a = (\text{number of inputs in } a)/p_a$. Note this means that the average number of inputs in a , E_a , depends on a 's input measure: Given a fixed number of spins in the region a , if p_a were greater, E_a would be smaller. (This is illustrated by comparison with 4(b), described below.) We can also consider a croupier's average number of inputs for the entire bubble, again defined as the number of inputs divided by the bubble's measure: $E_b = (\text{number of inputs in bubble } b)/p_b$.²⁸ A bubble-deviation for red over the red-black pair of velocity intervals in figure 9.4(a) is then the absolute difference between the height of the right-hand rectangle (the average number of inputs in the corresponding input region) and the average height over the entire bubble (the average number of inputs over the entire bubble) divided by the input probability of the bubble. Specifically, the bubble-deviation for red in a bubble b is defined, relative to an input probability measure, as the absolute value

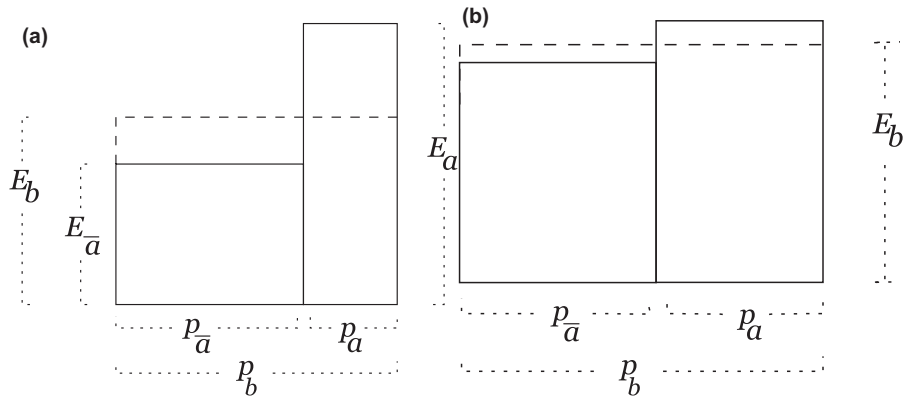


Figure 9.4 Initial condition distribution within a bubble. Width represents measure.

4(a): First input measure; 4(b): Second input measure

of the difference between the expectation E_b of the number of inputs conditional on b and the expected number E_a of inputs conditional on the set of inputs in a , divided by b 's measure. The bubble-deviation is equal to: $\left| \frac{E_b - E_a}{p_b} \right|$.

Figures 9.4(a) and (b) together illustrate how bubble-deviation depends on input measure. Both diagrams represent one croupier's distribution of spins over the same bubble: Numbers of spins in each region are the same: The area of the right-hand rectangle in (a) is equal to the area of the right-hand rectangle in (b), and similarly for the left-hand rectangles. However, because of the difference in input measures assigned to a (width in the two diagrams), (b) shows a smaller bubble-deviation than does (a): The difference between the height of the right-hand rectangle, and the height of the dashed line representing average number of inputs in the bubble is smaller in (b).

As mentioned above, according to the microconstancy inequality theorem, the smaller bubbles' measures are, and the smaller the bubble-deviations of a distribution are, the closer frequencies will be to the probabilities of outcomes defined by input probabilities. Figure (b) illustrates part of the idea: When an input measure sets probabilities so that it makes the (probability-weighted) average number of inputs in special subsets of bubbles close to the average number of inputs for the entire bubble (it makes bubble deviation small), the step-shaped curve of average numbers of inputs in a bubble will be close to flat. That means that the frequencies of outcomes in the subsets will be close to the sets' measures according to the given probability measure.

9.3.3. Mechanistic Probability

In order to describe mechanistic probability, it will help to extend the notation from the previous section slightly. Let A represent an outcome (e.g., red), and let a represent the set of initial conditions which can lead to A . Let b be a bubble in the input space—that is, a set containing possible inputs leading to both A and its negation

(or complement), \bar{A} . Then $P(a|b)$ is the probability of an input leading to A given that it occurs in bubble b . This probability is equal to $\frac{P(a \& b)}{P(b)}$, the probability that an input in bubble b would lead to outcome A divided by the probability that an input would fall in b .

So far we've simply assumed the existence of a probability measure that assigns values to subsets of the input space. Now we can work toward a way of defining an appropriate input measure. The following relationship will play a central role below. Given a set of collections of inputs to a bubbly device (e.g., a set of collections of spins by different croupiers), *if* there's an input measure that

1. Makes maximum bubble size small, and
2. Makes bubble-deviations small for all (most) of the collections of inputs (i.e., makes the collections "macroperiodic" (Strevens, 2003)), and
3. Makes probability $P(a|b)$ of A 's causes conditional on coming from b the same for all bubbles b (i.e., P is "microconstant" (Strevens, 2003)),

then the microconstancy inequality theorem implies that frequencies will be near probabilities in all (most) of the collections of inputs. The strategy for defining FFF mechanistic probability is to define an input measure which best satisfies conditions (1–3) relative to a special set of collections of actual initial conditions. If (1)–(3) are satisfied well enough by the resulting input measure—in the sense that bubble sizes (1) and bubble-deviations (2) are small for most collections of initial conditions, and (3) is true, then mechanistic probability will be well-defined and will exist—for example, for a particular actual wheel of fortune. The special set of collections of initial conditions is specified in the next few paragraphs (for other details see the appendix to this chapter and Abrams, forthcoming).

In the case of some causal map devices, such as a particular wheel of fortune, it happens that there is a large set of actual collections of inputs to similar devices, where each collection is produced by a single device—a croupier, in this case—over a single interval of time. More specifically, there is a large set of actual collections of inputs to causal map devices whose input spaces are similar to that of the wheel of fortune in question (figure 9.5). Some of these collections contain spins of a particular wheel of fortune by a particular croupier. Some contain spins of a roulette wheel by a croupier. Some of the collections may contain spins of old mechanical one-armed bandit slot machines, in which vertical wheels were caused to turn by the pull of a lever. Note that in each case, what matters is that there is a set of inputs to a device whose input space is similar to that of the particular wheel of fortune of interest to us. In the case of the roulette wheels, we can ignore the tossed balls, since there is nevertheless a spin of the wheel each time. Similarly, the fact that the wheels in the one-armed bandits are vertical doesn't change the fact that these wheels realize a type of causal map device whose input space consists of angular velocities within a certain general range. (For that matter, there are both vertically and horizontally mounted wheels of fortune.) Determining what all of these distributions are is impractical, though possible in principle.

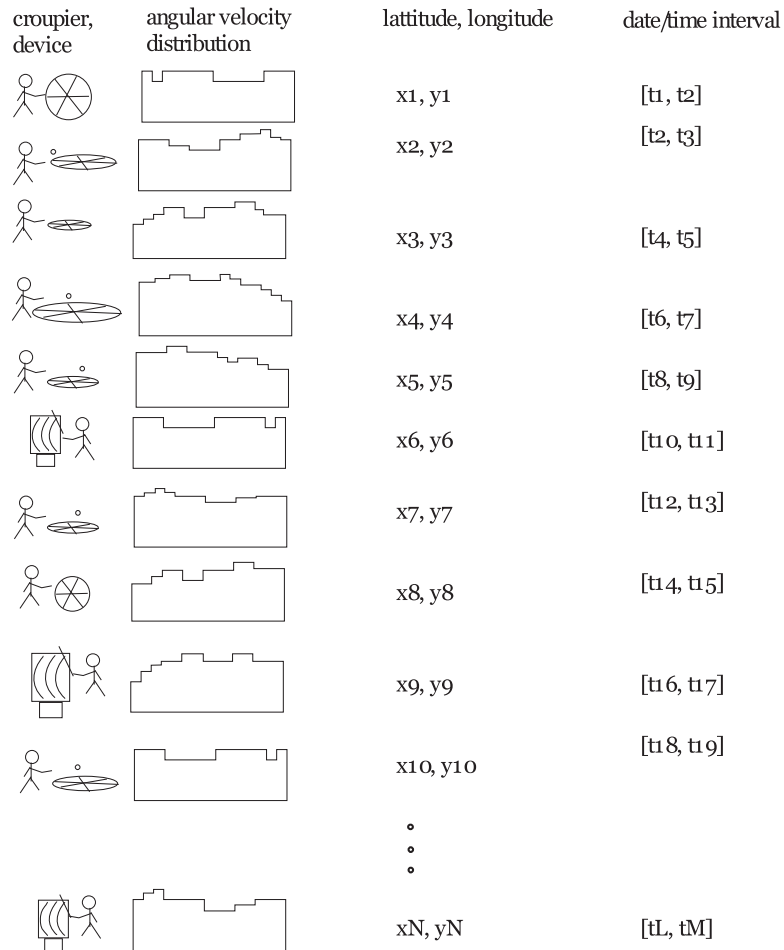


Figure 9.5 Distributions in a set of natural collections of inputs for one wheel of fortune, including spins of this wheel of fortune and others, roulette wheels, and mechanical slot machines

We can then define an input measure for a particular, concrete causal map device D (e.g., a wheel of fortune) relative to a large set of such “natural collections”, each containing a large number of inputs. These are those inputs to D and to other actual causal map devices D' which (a) exist in a large region of space and time around D , and (b) are similar to D in having roughly the same input space.²⁹ In addition, (c) a natural collection must consist of events that are all and only those produced by a single concrete device (e.g., a croupier) in a single interval of time. (This rules out gerrymandered collections containing only red spins, for example, and some other problematic cases; see Abrams, forthcoming).

The FFF mechanistic probability of an outcome (e.g., red) from a particular causal map device (e.g., an actual wheel of fortune) is then equal to the input probability of initial conditions which can lead to that outcome, where the input probability measure is one which:

- Makes the probability of the outcome (red) the same conditional on each member of a partition of inputs into bubbles ($P(A|b_i) = P(A|b_j)$ for all bubbles b_i and b_j), and
- Minimizes bubble sizes and bubble-deviations relative to all members of the natural collections of inputs.

FFF mechanistic probability then exists if those bubble sizes and bubble-deviations are sufficiently small.³⁰ Of course, since it's impractical to determine what all of these natural collections might be, it's impractical to completely verify that FFF mechanistic probability exists in a given case. Nevertheless, it's possible to get evidence for the existence of FFF mechanistic probability, and it's often reasonable to assume its existence. The dearth of adequate alternative explanations of stable frequencies in cases such as the wheel of fortune provides one reason for taking the postulation of FFF mechanistic probability to be reasonable. (See Abrams, forthcoming for further discussion.)

The definition of FFF mechanistic probability and the microconstancy inequality theorem have the following implications. (See (Abrams, forthcoming) for further discussion and defense of these points.)

- Frequencies in large sequences of trials will usually be close to mechanistic probabilities, because by definition, mechanistic probabilities are close to frequencies in most natural collections.
- Stable relative frequencies are explained by mechanistic probability. Bubbli-ness as it were compresses the diversity of frequencies in collections of initial conditions into the regularity of outcome frequencies.
- Bubbli-ness buffers against certain kinds of counterfactual variation in natural collections: Nearby worlds in which natural collections differ in small ways are ones in which frequencies of outcomes will generally be similar to what they are in the actual world. (Truthmakers for mechanistic probability are in actual world, however.)
- Since the same set of natural collections should apply to just about any causal map device with the same input space, probabilities of outcomes for causal map devices with the same input space and the same outcomes will differ in outcome probabilities primarily because of differences in their internal causal structure, as in the case of two wheels of fortune with different red and black wedge size ratios.
- A related point is that we can manipulate frequencies by modifying the structure of the device so that larger or smaller sets of initial conditions are mapped to outcomes—by changing sizes of wedges on wheel of fortune, for example. Thus mechanistic probability at least causally influences frequencies.

The last two points will be particularly important in the application of mechanistic probability to social contexts.

9.3.4. Causal Map Condition Categories

The following tale will help illustrate concepts which will be important for the application of mechanistic probability in social sciences.

At the end of a dirt road in a remote area there is an old, run-down casino, Casino Ruota, in which several wheels of fortune are often visible. Though the ratio between red and black wedge sizes is uniform within each wheel, the ratio is not the same for all wheels, which are otherwise identical. Thus red has a different probability on different wheels. Sal, the only croupier, races from one wheel to the next, spinning the wheel and quickly collecting and paying out at each wheel. She must be quick, because by magic that only Sal understands, each wheel comes into existence when it is spun, disappearing soon after it stops. (Sal also believes she is able to change the percentage of wheels favoring red and black using esoteric procedures she calls “policy interventions.”)

The fantasy that wheels magically come in and out of existence will help convey the point that the concepts described next don't depend on properties such as wheels' physical persistence. Since such properties will usually have no analogs in social science applications of mechanistic probability, it's important to see that they're essential to the strategy for applying FFF mechanistic probability explored in the rest of the chapter.

The properties of wheels of fortune in a collection such as Sal's can be classified into four *causal map device condition categories* (cf. Abrams, 2009, section 3.3.2):

1. The wheels have a set of mutually exclusive *outcomes*, in this case, red and black.
2. All wheels in the set share *common structuring conditions*: for example, having a certain diameter.
3. Wheels differ in which of several mutually exclusive *alternative structuring conditions* they involve. For example, for wheels in which the common structuring conditions require only two colors, red and black, with uniformly-sized wedges for each color on an evenly-weighted wheel, the alternative structuring conditions are different possible ratios between wedge sizes.
4. The wheels have a set of *circumstances of functioning*—conditions which change from trial to trial and change during the course of a trial: These include initial conditions (velocities), aspects of the process of spinning and slowing down, relation between the pointer and the wedge it points to, and so on.

Each trial—each spin—corresponds to a realization of the common structuring conditions, one consistent configuration from the set of alternative structuring conditions, and a configuration of initial conditions, which then cause the other circumstances of functioning to unfold until an outcome is produced. Call causal map devices that share common structuring conditions and outcomes, but have different alternative structuring conditions *alternative causal map devices* (relative to a

specification of such common and possible alternative structuring conditions). Classifying a set of similar causal map devices in this way is useful when we want to compare the effects of different alternative structuring conditions on probabilities of outcomes, given an assumption that the other aspects of the devices will be held fixed. As I'll explain below, this scheme will be useful in understanding the application of mechanistic probability in social science contexts.

9.4. APPLICATION IN SOCIAL SCIENCES

In this section I'll outline a way in which the preceding interpretation of probability can be applied in some social science contexts. I'll make repeated reference to the following illustration.

Hill et al. (1999) studied a sample of about eight hundred ten- to twelve-year-olds in Seattle during the 1990s. They identified several factors which seemed to affect frequency of an individual joining a gang between the ages of thirteen and eighteen. For example, the frequency of gang membership (0.233) among those whose families had household income in the lowest quartile was about twice as large (vs. 0.127) as for those with incomes in the highest quartile.

Family income itself of course does not guarantee any outcome, and following arguments in Section 9.2, I'll assume that lives of individuals in their physical and social context are deterministic. Thus given all of the details of an individual's circumstances, probabilities for events in a particular individual's life will be very close to 0 or 1. The relevant variables are too numerous and too fine-grained to measure in practice, but what happens to an individual is plausibly affected by details of her psychology, physiology, interactions with others (and their psychology and physiology), media broadcasts, weather, and much more. From this point of view, a portion of the world containing an individual has a certain very complex initial state at a time t_0 (e.g., tenth birthday), and then everything just unfolds—the machine runs. Along the way, certain identifiable outcomes (e.g., joining a gang) are realized: A state which incorporates those outcomes—those properties—becomes actual.

9.4.1. Some Say the Heart is Just Like a Wheel

Casino Sociale takes up large parts of the city of Seattle (and aspects of the rest of the planet). The gambling devices are both overlapping and ephemeral. They are realized by aspects of the city and its surroundings, each device focused on the life of a particular individual. No gamblers come from outside to play the games of chance, yet, routinely, a very complex "spin"—i.e., a configuration of initial conditions—is fed into the machine. A life unfolds; outcomes determined by that "spin" eventually occur or fail to do so. As before, the croupier is the same for each gambling device. However, here the croupier is realized by the same system that realizes the devices.

The heart isn't like a wheel of fortune; however, a heart along with the context in which it's embedded is.³¹ In the system studied by Hill et al. (1999), we can think of the life of a child, beginning from a specified time, as the focal element of a trial analogous to a spin of a wheel of fortune. However, since an outcome (e.g., gang membership by eighteen) is determined by the child's interactions with the world around him/her, the relevant causal map device includes much of what goes on in the child's nervous system and body, his/her neighborhood, the surrounding city, and to a lesser extent, events elsewhere on the planet. Call such a causal map device a "person-focused causal map device," or more briefly, a "personal causal map device." Not everything in the universe plays a role in such a device: Minute shifts in the molecular structure of a rock in Iowa usually won't affect children's lives, nor do subtle fluctuations in light reaching the earth from a distant pulsar. What does and does not count as part of the causal map device depends on the causal structure of the large system realizing it. (Note that "personal" is not meant to suggest that a personal causal map device belongs to or is associated with a particular individual; a personal causal map device is simply a device in which outcomes concern properties of the kind that persons realize individually.)

Different children are focal elements of different realizations of personal causal map devices. As with the ephemeral wheels of fortune in Casino Ruota, each trial—each course of operation of a device embedded in Seattle and its broader context—is a trial of a physically distinct device. Realizations of causal map devices relevant to Hill et al.'s (1999) study can be classified in terms of the four causal map device condition categories mentioned above. These devices have:

1. *Outcomes*: Properties such as joining or not joining a gang by a given age. (These are analogous to red and black outcomes for the wheel of fortune.)
2. A set of *common structuring conditions* shared by trials of the causal map devices, such as general properties of a person at age ten, general relations shared in virtue of living in Seattle, living in the United States in 1990. (These are analogous to the fact that it's a wheel of a certain diameter and mass which is spun, with given friction coefficients, and so on.)
3. A set of mutually exclusive *alternative structuring conditions*. Each personal causal map device realizes exactly one of these conditions. A set of mutually exclusive ranges of family income levels is one possible set of alternative structuring conditions. Racial and ethnic categories present a different set of alternative conditions. Conditions can be conjoined/intersected to produce a more fine-grained set of alternative conditions, such as these, for example: African American and high income, African American and low income, white and high income, white and low income.³² (The difference between income levels or racial categories corresponds to the difference between wheels of fortune with wedges of different sizes. For example, if we treat the red outcome as analogous to eventually joining a gang, then a wheel with large red wedges corresponds to living in a low-income family, while a wheel with smaller red wedges corresponds to living in a higher-income family.)

4. There are also *circumstances of functioning*, including a set of possible initial conditions at the beginning of a trial, along with everything else that varies in a child's life, family, neighborhood, the city, and so on—until, say, the child reaches the age of eighteen, at which point one of the gang member/not gang member outcomes has been realized. (Circumstances of functioning in the case of a wheel of fortune are simpler, including, primarily, only the initial angular velocity and the position of the wheel at each moment until it stops.)

Such conditions define a personal causal map type. A trial or realization of such a causal map type occurs whenever a set of common structuring conditions and one member of the set of alternative structuring conditions is realized, along with a specific set of initial conditions. All other properties involved in the realization of a personal causal map type count as circumstances of functioning, beginning from initial conditions. These include states of the focal person other than those included in the structuring conditions, states of other people in his/her community, physical configurations of the person's surroundings, and so on. Such trials take place repeatedly in Seattle and its broader context, just in Sal's Casino Ruota.³³

By deciding which conditions to study and estimate, a researcher implicitly selects a set of causal map types defined by conditions of the four kinds described above. For example, the researcher takes for granted certain background conditions, such as the fact that the only people being studied are those living in Seattle in the 1990s who are of a certain age. These are part of what define the common structuring conditions. The researcher also selects independent variables, which define alternative structuring conditions, and dependent variables, which define outcomes. Most circumstances of functioning usually go unmeasured, as they would be difficult to study in any detail.

However, the fact that researchers choose which aspects of a society to study doesn't imply that the causal relations connecting those aspects are themselves constituted by researchers' choices. Two researchers interested in different properties might (implicitly) select different sets of causal map types, which nevertheless might be realized by the exactly same persons in their physical and social contexts. What is a common structuring condition for one researcher (e.g., living in Seattle) might be an alternative structuring condition for another. (Suppose that city of residence is treated as an independent variable.) Moreover, what is a structuring condition for one researcher might be part of the circumstances of functioning for the other. (Suppose that individuals who relocate between cities during the period of study continue to be tracked. Note though that most circumstances of functioning will remain unmeasured by any researcher.³⁴) The two researchers *may* simply be asking about causal and probabilistic relationships between different properties that happen to be coinstantiated by the same system at the same time. On the other hand, suppose that two researchers choose to focus on the same outcomes and common structuring conditions, but on different sets of alternative structuring conditions. One researcher might focus on income while the other focuses on race.

These researchers are asking different questions: One asks about probabilities given various income properties, the other about probabilities given racial properties. Nothing I've said so far implies whether these questions have equally good answers, or whether one study might identify factors which are causally primary. Similarly, if one researcher takes for granted residence in Seattle and another doesn't, the first may miss interesting relationships between living in Seattle and the outcomes that interest her.

So far, I've explained how to see alternative personal causal map devices, analogous to wheels of fortune with different wedge sizes, within parts of a social system. However, I haven't yet argued that that other conditions required for the existence of mechanistic probability are satisfied by such causal map devices. In the next two sections I'll argue that it's plausible that many personal causal map devices are bubbly in a sense needed for mechanistic probability (in section 9.4.2) and that it's reasonable to see actual initial conditions which are inputs to personal causal map devices, as forming natural collections of inputs in the way needed for mechanistic probability (in section 9.4.3).

9.4.2. Life is Like a Box of Chocolates

My momma always said, "Life was like a box of chocolates. You never know what you're gonna get."—Forrest Gump³⁵

Individuals don't know what they're going to get, I'll argue, in this sense: For the kinds of outcomes which social scientists study—such as achievement of various kinds, well-being, drug use, voting behavior, participation in various groups—the complexity of the processes which actually govern a particular individual's life usually generates sensitive dependence of outcomes on initial conditions: Trivial differences in circumstances at one time routinely produce large differences in outcomes. Despite the fact that membership in a social category can make certain outcomes commonplace, it rarely guarantees any narrow set of outcomes. This point may seem obvious, but it will important to be clear about its implications here.

More importantly, I'll argue that it's plausible that the causal map devices realized by individuals in their social circumstances are typically bubbly: It's possible to partition the input space for a personal causal map device into many sets of similar circumstances, so that each set contains initial conditions leading to every outcome defined for the device. That is, it possible to partition the input space into bubbles.³⁶ A bubble of this kind will be a set of possible circumstances in individuals' lives—circumstances which probably all seem trivially similar to us—among which are variations, of an even more trivial character, which would lead to distinct social outcomes were they realized. (All of this is constrained by a choice of structuring conditions—see section 9.4.1). Individuals don't know what they're going to get, then, because they don't know into which portion of some bubble the initial conditions for their life fall. Causal paths between initial conditions and outcomes giving rise to the bubble structure are simply more complex than anyone can understand in practice.

Since rigorous demonstration of these claims would require impractically detailed empirical studies, my goal in this section is to give informal arguments to make it plausible that personal causal map devices are bubbly—that is, that bubbles of certain kinds exist—even if, as a practical matter, we can't usually identify particular bubbles. (As I noted at the end of section 9.1, the main goal of the chapter to make a certain hypothesis reasonably plausible. This is the claim that many probabilities of social outcomes are mechanistic probabilities. Given arguments in sections 9.1 and 9.2 that there is a need for such a hypothesis, and that no alternative seems to be in the offing, merely making mechanistic probability plausible for social science contexts is significant.)

At various points in the chapter I've suggested that the complexity of human lives makes it plausible that there is sensitive dependence of outcomes on details of circumstances at every stage of life. The story of the importance of my relative's pace and speed in putting on his coat (section 9.1) provided one illustration, and most people know of someone who got a job, chose their career path, or met a romantic partner as a result of what seemed to be a chance meeting between two people. Such events typically have other ramifications, both important and trivial. More prosaically, a delay in rounding a corner can lead to an encounter with a troubling or enlightening individual who would otherwise have been missed, which may in turn lead to missing a bus, being late for work, and so on. Or a slight puff of air in a small room can lead to an infection being transmitted. In a footnote to section 9.4.1, I mentioned that the circumstances of functioning for a causal map device can include circumstances such as who is in the same line in a shop, when a stop light changes, when a car turns into the street in front of a person, whether a new sidewalk fracture creates a puddle to be avoided, whether a strong wind slows one down, whether one is the target of a robbery which is or is not successful, whether a subtle insult is given at a moment when it is able to hit its target, vagaries of disease transmission and food nutrition, and so on. (Fiction sometimes uses small coincidences engineered by authors; that such events often seem uncontrived is some evidence of the ubiquity of small coincidences in real lives.³⁷) My claim here is that the complexity of social, biological, and physical interactions involved in human life means that even minor variations in factors like those just mentioned can ramify to produce divergent outcomes. The claim is not that human lives exhibit no robustness to minor variation in circumstances; I suspect that not all variations in circumstances would make a difference to outcomes studied by social scientists (cf. section 9.5.1). (On the other hand, since a bubble is any small set containing inputs leading to every outcome, if all minute variations in circumstances *did* make a difference to outcomes, and we knew it, justifying claims about mechanistic probability would be, if anything, easier, since very small sets would count as bubbles.) What I want to make plausible is only that in general, among sets of very similar circumstances, there are variations that would, if realized, produce each member of a wide range of outcomes that might be chosen for study within the social sciences. (Note that the very ubiquity of such sensitive dependence makes it hard to notice except in special cases. Part of the reason for

this is that the dependence on minor variation in circumstances in human lives is so common, and the ramifications of such variations so complex, that it would be impossible to trace the consequences of all such minor variations in practice.) If I'm right that in every small region of initial conditions for lives—that is, for a personal causal map device—there are small variations which lead to divergent outcomes, then person-focused social causal map devices are bubbly. That is, it would be possible to partition the input space of such a device into many sets, each containing some paths that lead to each outcome defined for the device.

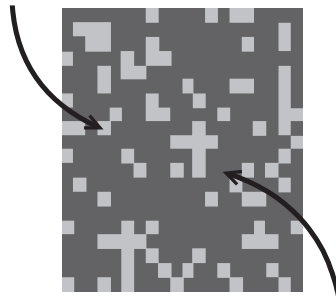
A detailed, concrete illustration of the bubblyness of a personal causal map device would be helpful, but would be impractical, requiring description in minute detail of numerous possible lives. Figure 9.6 provides a schematic representation, analogous to Figures 9.2(a) and (b), of a division of the input space of a personal causal map device into initial conditions leading to two outcomes: Gang membership (light gray) and lack thereof (dark gray). Similarity of conditions is represented by distance, probability is represented by area, and gang membership is less probable. Descriptions of what might ensue from two individual points in the input space are attached to the diagram. Though the stories in figure 9.6 are made up, they should not sound implausible. That they lack sufficient detail to predict the sequence of events conveys the extent to which further details of circumstances might have led to different outcomes than those in each story. Notice that the input space in figure 9.6 is represented in such a way that it's easy to divide it into many subsets of similar conditions containing initial conditions leading to both outcomes. That is, it's easy to show that this causal map device is bubbly. Of course I've designed this schematic illustration that way. One should imagine a much more fine-grained division with many more dimensions, however. The fact that points leading to gang membership or lack of gang membership are clustered in discernible regions is not essential to the diagram.³⁸

9.4.3. The Input Measure

In the last two sections I argued that we can view lives of persons in context as realizations of causal map devices (section 9.4.1), and that these causal map devices are usually bubbly (section 9.4.2). In this section I'll argue that it's reasonable to think that an input measure of the kind required by FFF mechanistic probability exists for these devices. That is, I'll argue that for many personal causal map devices, the social world distributes initial conditions across bubbles of the kind just described in a more or less even way. In particular, bubble-deviations will generally be very small relative to an input measure defined with respect to natural collections of those initial conditions. If this is correct, then the requirements for mechanistic probability would plausibly be met for many personal causal map devices.

First, as the story about Casino Sociale loosely suggests, the croupier for a social causal map device, for example, one focused on a child in Seattle, is the whole system in which that device is embedded. That is, different person-focused causal map devices, defined by different alternative structuring conditions (but identical

Child's mother struggles on and off with depression but works during some periods. Father has been around during some periods. Beginning at age 10, child and two younger siblings live with grandmother, who provides a friendly household with fairly strict rules. Child internalizes goal of doing well in school, and has some academic success while also doing regular babysitting work. Several childhood friends from immediate neighborhood have older siblings who join local gang, and those friends do so as well. Child remains committed to staying out of gangs, while also spending some time with friends who are gang members. However, child, along with friends, is mistreated in a few encounters with a local police officer, and child does poorly during one school year partly as a result of dealing with mother, who has been around but has been depressed. Father tries to spend more time with child in order to provide support but ends up leaving town for a job opportunity. Somewhat demoralized, child starts spending more time with gang friends, and ultimately joins gang after being beat up by members of a rival gang.



Focal child is from working-class, divorced family. Child sees both parents regularly. Parent in home is generally supportive but sometimes distracted by relationship issues. Some children harass child at school. This increases after child develops temporary limp due to playground fall, increasing tendency toward insecurity. Other parent is occasionally verbally abusive but also warm toward child. Close friend of child gradually increases use of marijuana over several years. Child sometimes threatened and robbed by a child who transfers to school. Some local police treat marijuana-using friend in heavy-handed way. Hangs out more with other drug-using friends who buy from gang members. Other gang members threaten child and friends are encouraged to join gang by peers; thinks about joining gang, but ends up encountering gang members less often after a close friend is expelled from school, and child gets more interested in basketball when another friend helps child improve skills, while one parent gets a less stressful job and then devotes more attention to child.

Figure 9.6 Schematic diagram of input space for personal causal map device for which gang membership (light gray) is less likely than avoiding it (dark gray). Events resulting from two points in input space are partially described. Distance represents similarity of initial conditions, and probability is represented by area.

input spaces, outcomes, and common structuring conditions) are realized within a larger device, the city of Seattle, or North America, and so on. This larger deterministic system is what chooses initial conditions to feed into each person-focused causal map device. (For example, these initial conditions might include the state of a child at age ten as well as her circumstances.) It may seem odd that the very

devices whose outcomes will end up having mechanistic probabilities, are themselves a part of the device that produces the inputs to each of these more limited devices. However, nothing in the logic of mechanistic probability rules out such a relationship between causal map devices and the devices that generate initial conditions to them. In this case there is a single, large natural collection of inputs, each fed into a different realization of a causal map device, and all produced by the single concrete device that is the larger system.

Second, note that the larger system which is the social croupier has been and will be spinning the wheels of many personal causal map devices, one for each individual in Seattle over a substantial period of time, at least.³⁹ The extremely large sequence of initial conditions fed into these devices can be divided into a large number of large subsequences—that is, a large set of natural collections of inputs, as required by the definition of FFF mechanistic probability.

Finally, it seems plausible that the croupier—society, Seattle, and so on—tends to distribute initial conditions in such a way that there are no *particular*, narrow, sets of conditions which are significantly more prevalent. It's implausible that the idiosyncratic, detailed circumstances that vary from individual trial to individual trial would do so in any systematic way that would produce large concentrations of inputs in small regions of the initial condition space. That is, initial conditions should be *very roughly* evenly distributed across bubbles and across subsets of bubbles leading to different outcomes. (A very roughly even distribution is all that's needed for FFF mechanistic probability—bubbliness then compresses the variation in the distribution of initial conditions into stable frequencies; cf. section 9.3.3.) Moreover, it's reasonable to think that all of this would hold for most subsequences of the very large sequences of initial conditions produced by Seattle the croupier.

Thus, for personal causal maps such as those implicit in Hill et al.'s study (1999), it seems plausible that there is a bubble-partition and an input measure that:

1. Makes the maximum input probability measure of all bubbles small (as described in section 9.4.2).
2. Makes the most of the natural collection of inputs macroperiodic, that is, such that bubble-deviations are small: Changes in input-probability-weighted average frequencies for gang membership, for example, and its complement within each bubble are small.
3. Makes the input probability measure P microconstant, that is, $P(a|b)$ of A 's causes conditional on b is the same for all bubbles b . (This requires adjusting the weights on portions of bubbles so that 1 and 2 aren't violated.)

Note that the set of actual initial conditions that defines the input measure for any one of the alternative personal causal map devices is the set of actual inputs to all of them, and the probability measure defined in terms of these inputs is the same for each alternative causal map device. The probabilities of outcomes for these devices differ only because of differences in their internal causal structures, defined by the different social conditions that they reflect.

As with the sensitive dependence of personal causal map devices, described above, illustrating the way in which initial conditions are distributed among bubbles for such devices would be difficult in practice, since it would require describing many actual situations in great detail. However, the more fine-grained the sensitive dependence is in a personal causal map device, the less plausible it seems that actual initial conditions would distribute themselves so as to favor any very small region in the input space containing outcomes leading to a single outcome, or that actual initial conditions would consistently favor one outcome over another across disparate portions of the input space. For example, it's implausible that initial conditions would consistently produce one of the precise scenarios described in figure 9.6 rather than others.

Thus it's reasonable to think that social outcomes defined for person-focused causal map devices such as those implicit in Hill et al.'s (1999) study have mechanistic probabilities: There is, plausibly, a bubble-partition and a probability measure over initial conditions to alternative causal map devices centered on ten-year-old Seattle residents (defined, e.g., by alternative parental income levels), such that the probability of gang membership conditional on each bubble is the same, such that every bubble's input measure is small, and such that the maximum bubble-deviation for the set of actual initial conditions over a long period of time is small.

If this is correct, then according to the microconstancy inequality theorem, when a particular causal map device, satisfying requirements for mechanistic probability, is repeatedly realized, frequencies of outcomes will usually be close to their mechanistic probabilities. These probabilities, I suggest, are often the probabilities, determined by the causal structure of social systems, that social scientists are in fact measuring, modeling, and estimating. (For example, on the present view, the frequency of joining a gang among children whose parents are in the lowest income quartile will probably be close to the mechanistic probability of gang membership for that condition.) More generally, I suggest that the reason that it can be useful to make predictions and develop policies based on social science research is that the system being studied is, underneath, a causal map device satisfying the requirements for FFF mechanistic probability.⁴⁰

I have argued only that all of the parts of this story about social probabilities are plausible, at the same time pointing out the severe practical difficulties of rigorously supporting the proposal. On the other hand, there does not seem to be any other account of social probabilities that can explain stable social frequencies. That does not make FFF mechanistic probability a purely theoretical postulate, however.⁴¹ Rather, FFF mechanistic probability is a complex explanatory hypothesis with plausibility, and moreover, the only one available. It may be that some future mechanistic probability interpretation other than FFF mechanistic probability—perhaps deriving from Strevens's (2011) or Rosenthal's (2010) work—will turn out to have more plausibility for social sciences. However, I'm unaware of any alternatives to mechanistic probability interpretations that could satisfy the desiderata specified above.⁴²

9.5. DISCUSSION

9.5.1. Independent Social Variables: Comparisons and Analogies

According to the present picture, the difference between values of independent variables—alternative structuring conditions such as high versus low family income—is like the difference between different wedge sizes for wheels of fortune: Differences in family income bias the frequencies produced (relative to a common input probability measure) by these alternative causal map devices. James Garbarino makes a related point about psychological factors which could be viewed as defining a set of alternative structuring conditions:

Most children are like dandelions; they thrive if given half a chance. Some are more like orchids. They do fine while young enough to be nurtured by loving parents, but wilt as adolescents subjected to peer competition, bullying and rejection, particularly in big high schools. (Garbarino, 1999, 51)

I would put the point this way: A personal causal map type defined by a more resilient, dandelion character biases probabilities—and frequencies—of better outcomes, compared to that defined by an orchid character. Some events that would send an orchid individual into a path that would (probably) lead to a bad outcome, would (probably) lead to good outcomes in many circumstances when a dandelion is the focal individual. (The instances of “probably” concern mechanistic probability.⁴³) In other words, compared to an orchid, a dandelion has larger—more probable, that is—neutral spaces (Wagner, 2005) leading to positive outcomes, where neutral spaces are regions of the input space in which outcomes are robust to possible small modifications of a path through life.⁴⁴ Similar points can be made about the effects of racial or economic differences and other factors on individual outcomes. (It’s plausible, however, that when there’s an alternative causal map device with larger neutral spaces for a given outcome, the neutral spaces are not so large as to destroy bubblieness. Otherwise there would be large regions of the initial condition space in which no variation, no coincidences, could lead to an outcome other than the one which is probable in that region.)

Note that despite the effects of alternative structuring conditions on aggregate effects, the outcome for any given individual is determined once the causal map device begins its operation, according to the present view. This point may be made clearer by considering an analogy with the motion of particles of dust in a liquid as they are affected by both Brownian motion and the underlying flow of the liquid. Although quantum mechanical indeterminism presumably affects these processes in real liquids, similar process would occur with deterministic molecular interactions, and simple models of Brownian reflect this assumption. On this view, a dust particle moves erratically because of variation in the numbers and forces of individual molecules impinging on it

deterministically. As a result, some dust particles may move against the flow for some time, but the average effect of the molecules buffeting dust particles will be such that average movement of particles is in the direction of the liquid's flow. The path of each dust particle is nevertheless completely determined by the particular impacts on it. On the present picture of social phenomena, even individuals who realize social conditions that generate an average "flow" toward outcomes with high mechanistic probability, may be those buffeted by life's deterministic "molecular" impacts so as to be pushed toward outcomes with low mechanistic probability.⁴⁵

9.5.2. Correlations in Initial Conditions

In the United States, race is correlated with income. This is probably due in part to the structure of causal map devices with different racial categories as alternative structuring conditions. For example, if a significant proportion of employers are more willing to hire people perceived as white than people perceived as black, FFF mechanistic probabilities of various incomes will be different for personal causal map devices corresponding to white as opposed to black focal individuals.

The correlation between race and income could also be due in part to correlations among initial conditions. For example, consider personal causal map devices whose input spaces consist of circumstances at the time of an individual's birth. If there is a correlation between race and income among parents—a correlation in initial conditions for children—and if there's a correlation between parents' income and the educational opportunities of children, the observed correlation between race and income may be partly the result of patterns in initial conditions, rather than the causal structure of personal causal map devices alone. Obviously, this correlation between race and parental income may be due to the causal structure of personal causal map devices focused on individuals in the parents' generation. Thus discrimination in one generation can cause a correlation between race and income among parents of individuals in the next generation, which affects the pattern of initial conditions for the following generation. This correlation in initial conditions can contribute to a similar race-income correlation in this second generation. Thus it could well be that to some extent, the race/income correlation is not merely regenerated by discrimination in every generation, but persists because the structure of current personal causal map devices as it were passes on race/income correlations from one generation to the next. This is in contrast to the picture painted above, in which alternative causal map devices generate different mechanistic probabilities solely due to their internal differences of causal structure *from the same input probability measure*, and frequencies are supposed to differ for the same reason. That is, the picture given above assumed that alternative causal map devices were in effect subject to the same overall patterns of inputs, and as a result could be treated as having the same input measure. In the race/income scenario just described, though,

frequencies differ in part because of differences in the ways in which initial conditions are distributed to causal map devices corresponding to black versus white children. It appears that personal causal maps for white children and for black children should be given different input measures. Mechanistic probabilities in this case should reflect these different input measures as well as differences in the internal causal structure of the alternative causal maps due, for example, to discrimination. Of course, the historical root of the correlation in initial conditions many generations ago undoubtedly has to do with discrimination as well, but the framework I've set up in this chapter doesn't have a clear role for the effects of discrimination which took place many generations ago.

Note, however, that if instead we define a set of alternative causal map devices in terms of race, parental income, and perhaps some other factors all at once, it can turn out that there are no remaining correlations in initial conditions that affect frequencies of outcomes differently in different alternative causal map devices. There may be more realizations of white plus high-income-parent causal map devices than black plus high-income-parent devices, at present, but that doesn't affect frequencies conditional on each device. I believe that with the right set of alternative structuring conditions, FFF mechanistic probability will apply in spite of correlations in initial conditions for less specific causal map devices.

9.6. CONCLUSION

Probabilities in social sciences might have all been artifacts of models, or might have literally referred to population frequencies. When frequencies are stable, however, there would seem to be something systematic about the underlying social system that produces that stability. Since such systematic facts would generate stable frequencies, it's reasonable to view them as playing a role in constituting objective probabilities inhering in the social system. Since we are interested not just in whatever frequencies in social systems happen to be, but also in manipulating these frequencies, it appears as if we do think that there are such probabilities. However, we have not had a good account of what it is about social systems that could count as this kind of probability; we have not had an interpretation of probability that can play the role of explaining stable frequencies. FFF mechanistic probability can play this role. The evidence for FFF mechanistic probability in social contexts is not as strong as one would like, however. I can only argue that it's plausible that certain conditions for the existence of FFF mechanistic probability are satisfied in social systems. However, given that there does not seem to be a good alternative, FFF mechanistic probability should be taken seriously as an account of the kind of probability to which many claims in social sciences implicitly refer.

APPENDIX: MECHANISTIC PROBABILITY: TECHNICAL ASPECTS

As mentioned above, the bubble-deviation of a bubble b is defined relative to an input probability measure as the absolute value of the difference between the expectation E_b of the number of inputs conditional on b and the expected number E_a of inputs conditional on the set of inputs leading to A (i.e. “ A ”) within b , divided by b ’s measure: $\left| \frac{E_b - E_a}{p_b} \right|$ (See figure 9.4(a)). The microconstancy inequality theorem A^1 (Abrams, forthcoming) says that the difference between the relative frequency of outcome A and the input measure of (the set initial conditions which lead to A) is constrained by bubble-deviations and bubble measures:

Theorem 1 *The difference between the relative frequency $R(A)$ of A outcomes and the probability $P(A)$ of A is less than the product of the maximum bubble-deviation S , the square of the maximum bubble size π , and the number of bubbles n :*

$$nS\pi^2 \geq |P(A) - R(A)|$$

(Abrams, forthcoming)

Mechanistic Probability exists if and only if (Abrams, forthcoming):

1. There’s a large set of natural collections (all inputs produced by single process in single interval of time) of inputs—a set containing all and only those collections of inputs to actual devices D^* such that:
 - a. The inputs are all and only those produced by a single physical device;
 - b. D^* has approximately the same input space as D ;
 - c. D^* occurs somewhere in a large spatiotemporal region around the location and time of D ;
2. (a) D is bubbly: There is a bubble partition for D with a large number of bubbles, and this bubble partition is such that:
 - b. A microconstant input measure (i.e., such that within-bubble outcome probabilities are uniform across bubbles), constructed through minimization of the sum of squares of bubble-deviations for all members of the set of natural collections, makes most of the collections macroperiodic relative to this input space and the device’s bubble partition. (Or: Most collections generate a small maximum bubble-deviation.)
 - c. Moderately expanding or contracting the spatial or temporal range across which natural collections are defined doesn’t affect whether condition is satisfied for outcome probabilities.

See (Abrams, forthcoming) for clarification and discussion of details.

ACKNOWLEDGMENTS

I'm grateful for Harold Kincaid's feedback, especially his help clarifying the presentation. (He's not responsible for the remaining unclarities.) Brenda Smith provided valuable discussion and suggestions concerning useful source material. I'm also grateful for helpful interactions with Aidan Lyon, David Teira, Don Ross, Erik Angner, Fred Chernoff, Jim Woodward, Julian Reiss, Ken Kollman, Lois Barnett, Mark Risjord, Norman Abrams, Philippe Huneman, Robert Northcott, Ron Mallon, Ted Benditt, as well as others who participated in workshops at which these ideas were presented: the conference on Current Issues in the Philosophy and Methodology of the Social Sciences, and the International Network for Economic Method Conference, both at the University of Alabama at Birmingham in 2010, and a presentation at the Université de Paris I in 2011. Finally, since this chapter grows directly out of the research which led to (Abrams, forthcoming), those acknowledged there are here, implicitly, as well. No one mentioned should be assumed to be in agreement with or even sympathetic to the ideas presented above.

NOTES

1. Frequency interpretations are widely taken to be the foundation of so-called frequentist statistical methods, but what these methods actually require is just an interpretation of probability that satisfies the assumptions of the methods. My view is that FFF mechanistic probability, described in this chapter, roughly satisfies those assumptions, and that it can be modeled by mathematical probabilities that precisely satisfy them.
2. Compare the popular advice to "follow your dreams, no matter what, and you will succeed."
3. Garfinkel (1981, 91ff) comments directly on the previously quoted passage from (Jencks et al. 1973).
4. Garfinkel (1981) repeatedly compares social processes to his view that a rabbit about to be eaten by a fox would nevertheless still have been eaten, had circumstances been slightly different. He takes this illustration as a good analogy for the correct way to think about many explanations in social sciences. I'll note that I think the story about the fox and the rabbit is implausible, especially if we consider subtle alterations in circumstances leading up to the moment of predation. Natural selection usually pushes organisms to the limits of their capabilities in one dimension or another; otherwise, there is room for selection to improve until some such limits are reached. Prey are often difficult to catch, and no doubt some prey are only just barely caught. In the usual highly complex environments, circumstances that can make a difference to outcomes are likely to be subtle, a point I will discuss elsewhere.
5. Garfinkel (1981) eventually seems to argue (in his chapter 6) that social science (and science in general) *can't* proceed where there is sensitive dependence of outcomes on subtle variations in initial conditions. What Garfinkel fails to emphasize is the extent to which social science only makes probabilistic claims about aggregate outcomes, which often do not exhibit pernicious sensitive dependence.
6. FFF mechanistic probability is closely related to other recently proposed interpretations of probability (Strevens, 2011; Rosenthal, 2010; Myrvold, 2012), but I believe it is more appropriate for understanding probabilities in social sciences.

7. Except, possibly, proposals similar to the one I describe here, as remarks I make below about Strevens's and Rosenthal's interpretations of probability suggest.

8. Sometimes methodological probabilities are alluded to without being explicitly described, as in "a total of 10,000 individuals were drawn from the national population database in a random procedure which was stratified by age, gender, and county of residence" (Wenzel, Øren, and Bakken 2008).

9. Bayesian statistical methods require that researchers choose a prior probability distribution or a class of such distributions. Is this an instance of social probability or methodological probability? I suggest the following: If the prior distribution is chosen arbitrarily, because it will wash out—because the choice of prior distribution has little influence on the inferences made in the end—then the priors are methodological probabilities. If, on the other hand, a prior distribution is tailored to facts about social systems, it may be that it is, at least in part, an estimate of a social probability.

10. The set of subsets must be closed under finite unions.

11. (3) is often replaced with a stronger requirement which generalizes it to infinite sums: $P(\bigcup_j^{\infty} A_j) = \sum_j P(A_j)$, if distinct A_j 's can have no members in common. This is usually paired with a requirement that the set of subsets is closed under infinite unions. Mechanistic probability will only satisfy the finite additivity requirement in (3) rather than this stronger requirement of countable additivity. It can nevertheless be mathematically convenient to model mechanistic probabilities with countably additive probabilities. Most mathematical treatments of probability as a measure require countable additivity, though if Ω is given only a finite number of subsets, finite additivity implies countable additivity.

12. See, e.g., Hájek, 2009b and Williamson, 2010 for other sets of criteria and further discussion.

13. There are ways to argue, concerning either scenario, that a rational subjective or epistemic probability assignment would make a frequency near .5 probable in this sense. However, common notions of subjective or epistemic probability won't capture the difference between these scenarios.

14. See, e.g., Gillies, 2000 and Hájek, 2009b for more thorough surveys of interpretations of probability and their advantages and disadvantages.

15. Known as a "reference class"—a class with reference to which frequencies are to be calculated.

16. Nor could frequencies in one set of events explain frequencies in another sets of events without additional, substantive assumptions (for example, that there are causal connections between the two sets of events).

17. This is true of Hoefer's (2007) theory despite his emphasis on what he calls statistical nomological machines (SNMs), such the wheel of fortunes I describe below. The problem is that Hoefer also defines chances in in term of frequencies alone when no such SNMs exist.

18. The notion of a propensity was originally proposed by Popper (1957). This notion is not related to propensity scores or any other statistical or methodological principles involving the same term.

19. The assumption that underlying causes are governed by probabilities is part of what's needed, for example, to justify a claim that errors in observation of a single physical state will have a Gaussian distribution.

20. It isn't difficult to modify this picture if some causes of behavior are indeterministic (cf. Strevens, 2003, chapter 2).

21. I believe that it's clearer how to apply FFF mechanistic probability in social sciences than it is to apply the other two interpretations, in part because FFF mechanistic probability was originally developed with the special sciences in mind, while Strevens's and

Rosenthal's interpretations apply more naturally in physical sciences. However, ideas developed in (Strevens, 2003, 2005) might be used to apply Strevens's or Rosenthal's interpretations in social sciences. I discuss the relationship between an aspect of Strevens's interpretation and mine in (Abrams, forthcoming).

22. The $^{-1}$ notation for the set of elements that is mapped to a given set is common in mathematics. The set A^{-1} is the result of the inverse of the function from inputs to outputs, somewhat in the way that applying $^{-1}$ to x produces, x^{-1} the inverse of x .

23. "Bubble" was suggested by the fact that a soap bubble reflects its surroundings in miniature; similarly, an input space bubble reflects the outcome space "in miniature".

24. In figure 9.2(b) the small rectangle in the input space is a bubble, in this case for a causal map device with three outcomes.

25. In addition to Strevens (2003), cf. Poincaré (1912, 1952 [1903], 1952 [1908]) and works by other authors on the method of arbitrary functions cited in von Plato (1994), as well as more recent work by Keller (1986) and Engel (1992).

26. Lebesgue measure generalizes summed lengths, areas or volumes to an arbitrary number of dimensions, while avoiding some technical problems with mathematical refinements that won't concern us here.

27. Recall that one of the axioms says that the probability of the union of two nonintersecting sets is the sum of their individual probabilities.

28. It makes sense to use average numbers of inputs, e.g., in the set a , rather than counting the croupier's numbers of inputs at each velocity in the input space, since it's likely that there is no more than one spin at any *particular* velocity, with no spins at all at most velocities. Tracking numbers of spins at particular velocities would make it difficult to see an overall pattern to a croupier's spins such as that illustrated in figure 9.3(b). It's such patterns that are the key to understanding mechanistic probability. Averaging over the input measure of a set in this way allows mechanistic probability to use an input measure that has been adjusted to reflect evidence of general tendencies to produce certain stable frequencies.

29. Other criteria might be required; see ~~below for further~~ discussion of this point for social science contexts.

30. See the appendix to this chapter and Abrams (forthcoming) for discussion the appropriate sense of minimization and what "sufficiently small" means. Abrams (forthcoming) discusses what is important concerning the number and size of the natural collections.

31. The title of this section is from Anna McGarrigle's song "Heart Like a Wheel" (McGarrigle, 2010).

32. Neither alternative structuring conditions nor common structuring conditions need correspond to what social scientists call "structural conditions."

33. Abrams, (2009) fleshes out some aspects of this picture in a biological context, using what I call "organism-environment history spaces." The framework described there is applicable to many social contexts with minimal modification.

34. These circumstances of functioning include, at various moments in time, e.g., who is in the same line in a shop, when a stop light changes, when a car turns into the street in front of a person, whether a new sidewalk fracture creates a puddle to be avoided, whether a strong wind slows one down, whether one is the target of a robbery which is or is not successful, whether a subtle insult is given at a moment when it is able to hit its target, vagaries of disease transmission and food nutrition, etc.

35. From the movie *Forrest Gump* (1994) directed by Robert Zemeckis, and written by Winston Groom and Eric Roth.

36. Technically, bubblieness is not independent of other requirements for mechanistic probability: The partition into bubbles must be one that allows satisfaction of the other conditions required for mechanistic probability. A consequence of this fact is that my use of “similar” in this paragraph is inessential, though useful for conveying the idea I have in mind. See appendix to this chapter and Abrams (forthcoming).

37. My favorite illustration is the 1998 film *Run, Lola, Run* by Tom Tykwer, in which differences between three counterfactual scenarios seem to depend in large part on small differences in timing in an initial sequence of events common to all three scenarios.

38. A similar diagram with intermixed gray and black dots would illustrate the idea just as well, if not better, but would make it difficult to see which outcome had greater probability.

39. A question remains about whether circumstances of ten-year-olds in other cities also should be viewed as helping to determine the input probability measure causal map devices relevant for Hill et al.’s study (1999). It may be that inputs to personal causal map devices for children in other cities, where the devices have the same input and output spaces, are relevant whenever patterns of correlations in initial conditions are similar to those in Seattle.

40. In some situations we are not only interested in probabilities and frequencies for trials of a given kind—e.g., of a given wheel of fortune or lives of children—but also in whether individual trials are probabilistically independent. Some statistical methods depend on such independence assumptions. To ask whether two trials or realizations of a single causal map device are independent is to ask a question about a complex device defined by two distinct trials of a simpler device. Thus, the wheel of fortune device discussed above is defined by an input space of angular velocities and an outcome space of red and black. We can also define a causal map device in which the input space consists of pairs of angular velocities for successive spins, and whose outcomes are pairs of red/black outcomes. In a more complex device such as this one, we can ask whether the outcome *A*-on-trial-one is independent of *A*-on-trial-two. Similarly, a question about independence of outcomes of two realizations of a given personal causal map device is really a question about a causal map device composed of two such devices. (In order to represent the outcomes of two trials together, we need to consider an outcome space that can represent both trials’ outcomes. We are able to make claims about probabilities of such joint outcomes only given probabilities of outcomes on one trial given the outcome of the other trial. This is exactly what an assumption of probabilistic independence of trials gives us, for example: We assume that the probability of *A* on one trial is equal to *A* probability on that trial given the outcome of the other. However, the outcome space of a simple causal map device doesn’t include joint outcomes. For example, the outcome space of a wheel of fortune device consists only of red and black, and FFF mechanistic probabilities for that device concern only those outcomes. Since mechanistic probabilities are defined only relative to a particular kind of causal map device, we have to consider a device whose outcome space includes joint outcomes of distinct trials in order to have mechanistic probabilities which give probabilities of one kind of outcome conditional on the other.) Questions about independence in a multi-trial causal map device in social contexts can be less clear than in the case of a wheel of fortune: Some realizations of the same personal causal map device may even involve direct interactions between the focal individuals. (Consider realizations corresponding to two children in the same family.) Nevertheless, there are reasons to think that bubblieness can often allow probabilistic independence even when there are causal interactions between realizations of different simple causal map devices. Strevens (2003, chapter 3) argues for this point.

41. An example of the latter would be Sober's (2005) "no-theory" theory of probability, which claims that objective probabilities are purely theoretical postulations; their theoretical role in particular sciences is all there is to no-theory probabilities.
42. Explaining why I think that Strevens's (2011) and Rosenthal's (2010) interpretations are unsuitable for probabilities in social sciences would require a long digression, but I'll mention that their interpretations depend on an input measure defined in terms of purely physical properties, which—I would argue—would usually be unsuitable as a basis for probabilities in social sciences. (Abrams [forthcoming, section 4.1] includes one general criticism of Strevens's [2011] way of defining an input measure.)
43. More specifically, I view the probabilities underlying the two instances of "probably" in this sentence as mechanistic probabilities conditional on a particular event type (e.g. on being an act of rejection), occurring during the functioning of possible realizations of a causal map type. See Abrams (2009) for a conception of an organism (e.g., person) in its environment consistent with the present conception of mechanistic probability, which allows a clear way of understanding such conditional probabilities.
44. See Wimsatt (2007) for related discussions of robustness in a variety of contexts.
45. William Upsi Wimsatt (1994) wrote that life as a white American is like riding a bicycle with the wind at one's back: It helps one along but its effects are typically unnoticed. The effects of the wind are harder to miss when one has to ride into it. My metaphor would fit Wimsatt's if air molecules were very large.

REFERENCES

- Abrams, Marshall. 2007. "Fitness and Propensity's Annulment?" *Biology and Philosophy* 22(1): 115–30.
- Abrams, Marshall. 2009. "Fitness 'Kinematics': Altruism, Biological Function, and Organism-Environment Histories." *Biology & Philosophy* 24(4): 487–504.
- Abrams, Marshall. Forthcoming. "Mechanistic Probability." In *Synthese*.
- Eagle, Antony. 2004. "Twenty-One Arguments Against Propensity Analyses of Probability." *Erkenntnis* 60(3): 371–416.
- Engel, Eduardo M. R. A. 1992. *A Road to Randomness in Physical Systems*. Berlin: Springer-Verlag.
- Garbarino, James. 1999. "Some Kids Are Orchids." *Time Magazine* 154(25): 51.
- Garfinkel, Alan. 1981. *Forms of Explanation: Rethinking Questions in Social Theory*. New Haven: Yale University Press.
- Gillies, Donald A. 2000. *Philosophical Theories of Probability*. New York: Routledge.
- Hájek, Alan. 1996. "'Mises Redux'-Redux: Fifteen Arguments Against Finite Frequentism." *Erkenntnis* 45(2–3): 209–27.
- Hájek, Alan. 2009a. "Fifteen Arguments Against Hypothetical Frequentism." *Erkenntnis* 70(2): 211–35.
- Hájek, Alan. 2009b. "Interpretations of Probability." In *The Stanford Encyclopedia of Philosophy*, Spring 2009 ed., Edward N. Zalta, ed. Available at <http://plato.stanford.edu/archives/spr2009/entries/probability-interpret/>.
- Hill, Karl G., C., Howell James, Hawkins, J. David, and Battin-Pearson, Sara R. 1999. "Childhood Risk Factors for Adolescent Gang Membership: Results from the Seattle Social Development Project." *Journal of Research on Crime and Delinquency* 36(3): 300–22.

- Hofer, Carl. 2007. "The Third Way on Objective Probability: A Sceptic's Guide to Objective Chance." *Mind* 116(463): 449–596.
- Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., and Michelson, S. 1973 [1972]. *Inequality: A Reassessment of the Effect of Family and Schooling in America*. New York: Harper Colophon.
- Keller, Joseph B. 1986. "The Probability of Heads." *The American Mathematical Monthly* 93(3): 191–97.
- Lewis, David. 1980. "A Subjectivist's Guide to Objective Chance." In *Studies in Inductive Logic and Probability*, vol. II, Richard C. Jeffrey, ed. Berkeley: University of California Press. Reprinted in Lewis, 1986.
- Lewis, David. 1986. *Philosophical Papers*, volume II. Oxford: Oxford University Press.
- Lewis, David. 1994. "Humean Supervenience Debugged." *Mind* 103(412): 473–90.
- Loewer, Barry. 2001. "Determinism and Chance." *Studies in the History and Philosophy of Modern Physics* 32B(4): 609–20.
- Loewer, Barry. 2004. "David Lewis's Humean Theory of Objective Chance." *Philosophy of Science* 71(5): 1115–125.
- McGarrigle, Anna. 2010. "Heart Like a Wheel." Lyrics of the song "Heart Like a Wheel," published by Garden Court Music (ASCAP). www.mcgarrigles.com/music/kate-and-anna-mcgarrigle/heart-like-a-wheel.
- Millikan, Ruth Garrett. 2000. "What Has Natural Information to Do with Intentional Representation?" In her *On Clear and Confused Ideas*, 217–37. Cambridge: Cambridge University Press.
- Mofitt, Robert. 2008. "A Primer on U.S. Welfare Reform." *Focus* 26(1): 15–25.
- Myrvold, Wayne C. 2012. "Deterministic Laws and Epistemic Chances." In *Probability in Physics*, Meir Hemmo and Yemima ben Menahem, eds. Berlin: Springer-Verlag.
- Poincaré, Henri. 1912 [1896]. *Calcul des Probabilités*, 2nd ed. Paris: Gauthier-Villars.
- Poincaré, Henri. 1952 [1903]. "The Calculus of Probabilities." Chapter XI in *Science and Hypothesis*. In Poincaré 2001.
- Poincaré, Henri. 1952 [1908]. "Chance." Chapter 4, Part I in *Science and Method*. In Poincaré 2001.
- Poincaré, Henri. 2001. *The Value of Science*. New York: The Modern Library.
- Popper, Karl R. 1957. "The Propensity Interpretation of the Calculus of Probability, and the Quantum Theory." In *Observation and Interpretation*, S. Körner, ed., 65–70. London: Academic Press.
- Rosenthal, Jacob. 2010. "The Natural-Range Conception of Probability." In *Time, Chance, and Reduction: Philosophical Aspects of Statistical Mechanics*, Gerhard Ernst and Andreas Hüttemann, eds., 71–90. Cambridge: Cambridge University Press.
- Sober, Elliott. 2005. "Evolutionary Theory and the Reality of Macro Probabilities." In *The Place of Probability in Science*, Ellery Eells and James Fetzer, eds., 133–61. Berlin: Springer.
- Strevens, Michael. 2003. *Bigger Than Chaos: Understanding Complexity through Probability*. Cambridge, MA: Harvard University Press.
- Strevens, Michael. 2005. "How Are the Sciences of Complex Systems Possible?" *Philosophy of Science* 72(4): 531–56.
- Strevens, Michael. 2011. "Probability Out of Determinism." In *Probabilities in Physics*, Claus Beisbart and Stephann Hartmann, eds., pp. 339–64. Oxford: Oxford University Press.
- Von Plato, Jan. 1994. *Creating Modern Probability*. Cambridge: Cambridge University Press.
- Wagner, Andreas. 2005. *Robustness and Evolvability in Living Systems*. Princeton, NJ: Princeton University Press.

- Wenzel, Hanne Gro, Øren, Anita, and Bakken, Inger Johanne. 2008. "Gambling Problems in the Family—A Stratified Probability Sample Study of Prevalence and Reported Consequences." *BMC Public Health* 8(412). Available at <http://www.biomedcentral.com/1471-2458/8/412>, doi:10.1186/1471-2458-8-412.
- Williamson, Jon. 2010. *In Defence of Objective Bayesianism*. Oxford: Oxford University Press.
- Wimsatt, William C. 2007. *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge, MA: Harvard University Press.
- Wimsatt, William Upski. 1994. *Bomb the Suburbs*, 2nd ed. New York: Subway and Elevated Press.