Ellery Eells, James H. Fetzer *Editors*

VOLUME 284 BOSTON STUDIES IN THE PHILOSOPHY OF SCIENCE

The Place of Probability in Science

In Honor of Ellery Eells (1953–2006)

Springer

THE PLACE OF PROBABILITY IN SCIENCE

BOSTON STUDIES IN THE PHILOSOPHY OF SCIENCE

VOLUME 284

For further volumes:
http://www.springer.com/series/5710

THE PLACE OF PROBABILITY IN SCIENCE

In Honor of Ellery Eells (1953–2006)

Ellery Eells • J.H. Fetzer
Editors

Springer

Editors
Ellery Eells
Springer
3311 GC Dordrecht
Netherlands

Prof. J.H. Fetzer
University of Minnesota
Dept. Philosophy
Duluth MN 55812
USA
jfetzer@d.umn.edu

ISBN 978-90-481-3614-8 e-ISBN 978-90-481-3615-5
DOI 10.1007/978-90-481-3615-5
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2010925023

Cover design: Boekhorst Design b.v.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Ellery Eells
in memoriam

Ellery Eells (1953–2006)

Preface

Science aims at the discovery of general principles of special kinds that are applicable for the explanation and prediction of the phenomena of the world in the form of theories and laws. When the phenomena themselves happen to be general, the principles involved assume the form of theories; and when they are particular, they assume the form of general laws. Theories themselves are sets of laws and definitions that apply to a common domain, which makes laws indispensable to science. Understanding science thus depends upon understanding the nature of theories and laws, the logical structure of explanations and predictions based upon them, and the principles of inference and decision that apply to theories and laws. Laws and theories can differ in their form as well as in their content. The laws of quantum mechanics are indeterministic (or probabilistic), for example, while those of classical mechanics are deterministic (or universal) instead. The history of science reflects an increasing role for probabilities as properties of the world but also as measures of evidential support and as degrees of subjective belief. Our purpose is to clarify and illuminate the place of probability in science.

The fundamental conceptions of probability that matter to science are both objective as properties of the world: *the frequency conception*, for which "probabilities" stand for relative (or limiting frequencies) of outcomes (such as heads) across sequences of trials (such as tosses of coins); and *the propensity conception*, for which "probabilities" stand for the strength of the causal tendencies for specific trials (or sequences of trials) to bring about specific outcomes. The frequency conception is a collective concept that applies to sequences of trials collectively (as a group). There is no relative frequency (or only a derivative value) for an outcome to occur on a single trial, although they are sometimes described that way. By comparison, the propensity conception is a distributive concept that applies to trials distributively (one by one). There are causal propensities for outcomes to occur on single trials, where different members of a sequence of trials can have causal propensities that vary from one trial to another when they occur under different conditions.

Evolutionary theory and quantum mechanics are among the most important scientific contexts in which objective probabilities play a role. The notions of biological fitness and of radioactive half-life, for example, both appear to be probabilistic properties. This volume includes several chapters devoted to their clarification. And it is quite important to distinguish between subjective conceptions of probabilities as properties as *beliefs about probabilities* (say, believing that the probability of heads on the next toss with a coin equals 1/2) and as *degrees of belief* (say, having a degree of belief equal to 1/2 that the next toss will be heads). Beliefs about probabilities are true when the world has those properties; degrees of belief are true when people do. These properties are closely related, since it is widely assumed that beliefs about probabilities determine degrees of belief. David Lewis even calls this relationship "the Principal Principle". It sounds simple, but understanding how it works is complex.

We begin with a comprehensive introduction to alternative conceptions of objective probability and the difficulties that they confront. The first section then follows with three studies of special problems that arise within this context and the comparative merits of different accounts. The second section addresses the nature of lawfulness and of relations between micro- and macro-probabilities, especially with reference to the concept of fitness in evolution. The third section confronts some of the difficulties confronted by causal conceptions of probability, especially within the quantum domain. The fourth extends the discussion to principles of inference and decision. The last chapter relates propensities and frequencies to the framework of inference to the best explanation. It all begins with an introduction that integrates these contributions and their interconnections, and it ends with an expression of our gratitude to the distinguished contributors who have made this collection possible.

When I retired from the University of Minnesota Duluth in June 2006 after 35 years of college teaching and moved to Madison, my wife, Jan, and I were looking forward to spending time with Ellery and his wife, Joanne, with whom we had spent many enjoyable evenings together over the years. We met in 1980 while I was visiting The University of North Carolina at Chapel Hill and Ellery was teaching at North Carolina State. We had only just moved into our new home in Oregon, WI, when we learned that Ellery had been hospitalized. Jan and I were able to visit him at Meriter Hospital shortly before he died on August 10. It was an acute loss for me and for the fields of philosophy to which he so richly contributed. Fortunately, we had completed our work on this project, including the Introduction of which he was the principal author. I have done my best to preserve his words throughout and offer this work as a monument to the excellence of his intellect as well as of his life as a husband, father, teacher and scholar.

James H. Fetzer

Contents

Contributors

Martin Barrett earned an M.S. in mathematics at the University of Michigan, and from 1980 to 1985 worked in scientific programming as an original contributor to the design and development of the McIDAS system for studying meteorological data from geostationary satellites. Currently a graduate student in philosophy at the University of Wisconsin at Madison, his interests lie principally in the philosophy of science, but extend to Bayesian epistemology, decision theory, and logic. Barrett has co-authored papers with several of the contributors to this anthology on causality, probabilistic epistemology, and biology. With mathematician Alan Macdonald, he has published an analysis of the fundamental equation of thermodynamics in *The American Journal of Physics*.

Ellery Eells, Professor of Philosophy at the University of Wisconsin at Madison (now deceased), earned undergraduate degrees in mathematics and in philosophy at the University of California at Santa Barbara and received his Ph.D. in philosophy at the University of California at Berkeley in 1980. His areas of interest and research include general philosophy of science, rational decision theory, causality, probability, and inductive logic. His publications include *Rational Decision and Causality* (Cambridge 1982), *Probabilistic Causality* (Cambridge 1991), and numerous articles on rational decision, causation, probability, and inductive logic. His final publication, most appropriately, is *The Place of Probability in Science*, which he co-edited with James H. Fetzer.

James H. Fetzer, McKnight University Professor Emertius at the University of Minnesota, Duluth, earned his Ph.D. in the history and philosophy of science at Indiana University in 1970. He has published more than 20 books in the philosophy of science and on the theoretical foundations of computer science, artificial intelligence, and cognitive science, including *Scientific Knowledge* (D. Reidel 1981), *AI: Its Scope and Limits* (Kluwer 1990), *Philosophy of Science* (Paragon 1993), *Philosophy and Cognitive Science*, 2nd edn (Paragon 1996), and *Computers and Cognition* (Kluwer 2001). He is founding editor of the journal *Minds and Machines* (Kluwer) and of the professional library *Studies in Cognitive Systems* (Kluwer). Two of his latest books, *The Evolution of Intelligence* (2005) and *Render Unto Darwin* (2007), appeared from Open Court.

Branden Fitelson is an Associate Professor in the philosophy department at UC-Berkeley. He also is a member of the Group in Logic and the Methodology of Science and the Cognitive Science Core Faculty at Berkeley. Branden has held non-teaching positions at NASA's Goddard Space Flight Center and (more recently) at Argonne National Laboratory. Before joining the philosophy department at Berkeley, Branden held teaching positions (in Philosophy) at the University of Wisconsin at Madison, the University of Illinois at Urbana-Champaign, Stanford University, and San Jose State University. His current research interests include the philosophical foundations of induction and probability, automated reasoning – which used to be called "automated theorem proving" – and the psychology of inference.

Malcolm R. Forster was educated in applied mathematics and philosophy at the University of Otago, New Zealand. He earned his Ph.D. in philosophy at the University of Western Ontario in Canada and held a Postdoctoral Fellowship in the Department of Mathematics and Statistics at Monash University, Australia. He is currently Professor of Philosophy at the University of Wisconsin at Madison, where his research interests include the philosophy of physics, the philosophy of biology, foundations of statistics, William Whewell's methodology of science, and cognitive science His articles have appeared in *Philosophy of Science*, *The British Journal for the Philosophy of Science*, and *Studies in the History and Philosophy of Science*, among other journals.

Daniel L. Hausman was educated at Harvard, New York University, Cambridge University and Columbia University, where he received his Ph.D. in 1978. Currently Herbert A. Simon Professor of Philosophy at the University of Wisconsin at Madison, he is the co-founder and past co-editor of the international journal, *Economics and Philosophy*. His books include *Capital, Profits and Prices* (Columbia 1981), *The Inexact and Separate Science of Economics* (Cambridge 1992), and *Essays on Philosophy and Economic Methodology* (Cambridge 1992). His interest in economics and methodology has led to extensive and continuing research regarding the nature of causation, which culminated in the publication of *Causal Asymmetries* (Cambridge 1998). He is also co-author (with Michael McPherson) of *Economic Analysis, Moral Philosophy, and Public Policy* (Cambridge 2006).

James Hawthorne, Associate Professor of Philosophy at the University of Oklahoma, received his Ph.D. from the University of Minnesota. He served as a research scientist in artificial science at the Honeywell Systems Research Center in Minneapolis for 9 years. His research interests focus upon inductive, probabilistic, and nonmonotonic logics and related philosophical issues in logic, mathematics, and the sciences. He has published articles for the *Stanford Encyclopedia of Philosophy*, *Synthese*, *Mind*, and the *Journal of Philosophical Logic* on a wide range of subjects, including inductive logic, Bayesian confirmation, the commutativity of updates based on uncertain evidence, the Bell inequality and nonlocality in quantum theory, mathematical instrumentalism, and nonmonotonic conditionals.

Robin Pemantle is a professor in the Department of Mathematics at the University of Pennsylvania in Philadelphia. He earned his Ph.D. in probability theory at the Massachusetts Institute of Technology in 1988. He has held positions at Ohio State University, the University of Wisconsin at Madison, and Oregon State University, as well as research appointments at the University of California at Berkeley, Cornell University, and Oregon State University. He has published extensively in the area of mathematical probability theory and has interests in the social sciences and in the education of children in mathematics. Within probability theory, his main interests include tree-index processes and branching processes as well as Brownian motion.

Brian Skyrms is the Distinguished Professor of Social Science, Logic and Philosophy of Science, in the School of Social Science at the University of California at Irvine and Professor of Philosophy at Stanford. He earned his Ph.D. at the University of Pittsburgh, and his academic distinctions include membership in The National Academy of Science and in The American Academy of Arts and Sciences, being both a Guggenheim Fellow and Fellow at the Center for Advanced Study in the Behavioral Sciences, and receiving the Philosophy of Science Association's Lakatos Prize. His publications include *Causal Necessity* (Yale 1980), *Pragmatics and Empiricism* (Yale 1984), *The Dynamics of Rational Deliberation* (Harvard 1990), *Choice and Chance*, 4th edn (Wadsworth 2000), *Evolution of the Social Contract* and *The Stag Hunt and the Evolution of Social Structure* (both Cambridge, 2004), and *Signals: Evolution, Learning and Information* (Oxford, 2010).

Elliott Sober is the Hans Reichenbach Professor of Philosophy and a William Vilas Research Professor at the University of Wisconsin at Madison, where he has taught since 1974. His research is in philosophy of science, especially in the philosophy of evolutionary biology. His books include *The Nature of Selection – Evolutionary Theory in Philosophical Focus* (MIT 1984), *Reconstructing the Past – Parsimony, Evolution, and Inference* (MIT 1988), *Philosophy of Biology* (Westview 1993), *From a Biological Point of View – Essays in Evolutionary Philosophy* (Cambridge 1994), and (with David Sloan Wilson) *Unto Others – The Evolution and Psychology of Unselfish Behavior* (Harvard 1998). His most recent book is *Evidence and Evolution – The logic Behind the Science* (Cambridge 2008).

Wolfgang Spohn studied philosophy, philosophy of science, logic, and mathematics at the University of Munich. He earned his Ph.D. with a thesis on the foundations of decision theory and his Habilitation with a thesis on the theory of causation. After an associate professorship at Regensburg and full professorship at Bielefeld, he currently holds a chair for philosophy and philosophy of science at the University of Konstanz since 1996. He served as Editor-in-Chief of *Erkenntnis* from 1988–2001 and as the director of the DFG research group "Logic in Philosophy" from 1997–2003. His many papers deal with epistemology, induction and probability, causation, explanation, and general philosophy of science, philosophical logic, philosophy of language and mind, decision theory, game theory, and practical rationality. Some are collected in *Causation, Coherence, and Concepts* (Springer 2008).

Paul Weirich, Professor of Philosophy at the University of Missouri at Columbia, earned his Ph.D. at UCLA. He is past chair of his department, and previously held a faculty position at the University of Rochester. As a graduate student, his interest in probability prompted a dissertation treating subjective probability in decision theory. He has written four books on related topics: *Equilibrium and Rationality: Game Theory Revised by Decision Rules* (Cambridge 1998), *Decision Space: Multidimensional Utility Analysis* (Cambridge 2001), *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances* (Oxford 2004), and *Collective Rationality* (Oxford 2009). His current research includes both decision theory and game theory.

Introduction

The purpose of philosophy, broadly conceived, can be described as that of attempting to resolve heretofore unsolved conceptual and theoretical problems, especially those that lie at the foundations of knowledge and values. There is an intimate relationship between philosophy and language, since most philosophical problems arise as a result of ambiguities and vagaries in the use of language related to knowledge and values, such as the meaning of "belief", "justification", and "truth", in the case of knowledge, and the meaning of "good", "right", and "fair", in the case of values. Given the central role of science in acquiring reliable knowledge, the study of science has become of special importance within philosophy. And within philosophy of science, in turn, no other concept possesses the centrality and importance as does that of probability.

The crucial importance of probability within science can be illustrated in relation to the concept of a probabilistic explanation, which plays a central role in the natural sciences, the social sciences, and – arguably – even the humanities, at least when the discipline of history is subsumed thereby. Few historians would claim that what we know of history is "certain" rather than merely "probable"; and only the rash would hazard to project future trends on the basis of present knowledge with anything more than "probabilistic confidence". The links that can relate the future to the past take the form of statistical correlations, on the one hand, and laws of nature, on the other. Mere correlations, however, are not especially reliable and can change across time. Because laws cannot be violated and cannot be changed, they afford a more secure foundation for systematic predictions, and laws are indispensable for explanations.

Thus, a "covering-law" schema for probabilistic explanations, where probabilistic laws subsume specific cases when their conditional antecedents correspond to the initial conditions that were present in those specific instances, assumes the following form:

Schema S:	Covering Law	P(B/A) = p	Explanans
	Initial Conditions	Ac	
		==== [p]	
	Event to Explain	Bc	Explanandum

Here, the explanatory premises (known as "the explanans") include a covering law of the form, "P(B/A) = p", which asserts the probability P of B, given A, is equal to p, and description of initial conditions by virtue of which that law applies to this case, which in this instance is "Ac" ("c is A"), where the conclusion that is to be explained (known as "the explanandum") describes the relevant outcome, that is "Bc" ("c is B"). The link between the explanans as premise and the explanandum as conclusion is given by the double-line, which indicates an inductive relationship, and "[p]", which indicates the (probabilistic) logical strength with which the explanans subsumes the explanandum.

Simple examples can be drawn from games of chance, such as flips of coins, throws of dice, and draws of cards. Suppose, for example, that we assume that, on a single throw with a cubical, six-sided die, the probability of each of its sides – ace, deuce, trey, and so forth – coming up equals 1/6. The probability of some member of the set coming up is equal to the sum of their values, which is 1/6 times itself six times, equaling 1. The probability of one or another in some subset, such as an ace or a deuce, then turns out to be equal to their sum, which is 1/6 plus 1/6, or 1/3. And the probability of getting, say, an ace on one toss and another ace on a second toss is equal to the product of their probabilities, which would be 1/6 times 1/6, or 1/36. Then suppose that we wanted to explain why a specific cubical, six-sided die d of this kind D came up showing an ace A, given a throw T. Its explanation could assume the following form as an instance of S:

Example:	Covering Law	P(A/D & T) = 1/6	Explanans
	Initial Conditions	Dd & Td	
		= = = = = = [1/6]	
	Event to Explain	Ad	Explanandum

Here, the outcome "Ad" of showing an ace is explained on the basis of a covering law asserting that, for a die d of kind D that is tossed T, the probability P of an outcome of kind A equals 1/6. Interestingly, since this outcome has only a low probability given the relevant initial conditions and probabilistic covering law, this is not an outcome that we would be inclined to predict. On the contrary, the occurrence of a non-ace is equal to 5/6, which means that the outcome of a non-ace has a high probability and would be predictable. Even in the case of outcomes of low probability, arguments of this form still provide adequate explanations when they explain them. But they raise questions as to whether their explanandum-sentences should be called "conclusions".

In the case of games of chance, it is tempting to suppose that we know what we mean by the use of the term "probability" in describing situations of this kind, because we tend to assume that their values are equal to the number of possible outcomes (ace, deuce, trey, and so forth) divided by the number of outcomes of interest (an ace in this case). So the probability equals 1/6. Indeed, this interpretation of the meaning of probability is known as "the classic interpretation". But its tenability depends on the tacit assumption that the possible outcomes are equally probable. What if a coin were bent (a die were loaded, a deck were stacked)? Then those possible outcomes would not be equally probable and the classic interpretation would not apply. What should be done then? What would it mean to ascribe a "probability" in those cases?

The most important alternatives for understanding probabilistic covering laws are the frequency and the propensity accounts, which are discussed in most of the chapters of this volume. The first construes probabilities as relative frequencies, the second as causal tendencies. While that is the central problem addressed here, many others are closely related, including the connection between the value of "p" in covering laws and of "[p]" as the strength of the link between explanans and explanandum. And other important ramifications are investigated here, including the role of probabilities of either kind for drawing inferences and making decisions as well as special problems for probabilities that arise within contexts that are as diverse as evolutionary theory and quantum mechanics. Surveying these problems from diverse philosophical points of view, the present studies contribute to clarifying the place of probability in science.

Prologue

Objective Probability Theory Theory

Ellery Eells

Among the most subtle aspects of any philosophical investigation is settling upon the criteria of adequacy that provide a standard or a measure for evaluating alternative proposals and recommendations. Thus, Carl G. Hempel has suggested that there are three primary desiderata that matter in explicating the meaning of words, phrases, or expressions, namely: (1) that the syntax of those linguistic entities must be rendered explicit (as one-place, two-place, or more-place predicates, for example); (2) that the semantics of those syntactic entities must be relevant to the contexts of their usage (where concepts that fit one context may be inadequate for another); and (3) that the combination of that semantics with that syntax succeeds in attaining the pragmatic objective of clarifying and illuminating the meaning of that linguistic entity within that context of usage (Hempel 1952). Here we shall refer to these desiderata as the conditions of syntactic determinacy, semantic relevance, and pragmatic significance.

Ellery Eells pursues these objectives in his discussion of the standards that might be required, suggested, or urged for the adequacy of philosophical theories of objective probability. Probabilities are "objective" when they are supposed to be properties of the world as opposed to our beliefs about the world, which makes them "subjective". As he observes, various theories or interpretations of probability – characteristically in relation to specific mathematical functions that obey particular axioms, including principles of summation and of multiplication – appear in the philosophical and the statistical literature. Bayesian theorists, for example, have suggested the usefulness, in some contexts at least, of interpreting probability subjectively or as a measure of some particular agent's "degrees of belief", which may differ from person to person at the same time and even for one person at two different times. Others, of course, have proposed various objective interpretations of probability, including especially varieties of "frequency" and of "propensity" conceptions as properties of the world.

The double occurrence of the word "theory" in the title of Eells' paper is of course deliberate. Like Hempel, he is concerned with the conditions of adequacy that an adequate explication of probability must satisfy, especially if it happens to be an explication of probability as a physical magnitude. He implements the conditions of syntactical determinacy by adopting Wesley Salmon's requirement of admissibility (Salmon 1967), which Eells' takes to entail satisfying some standard axiomatization of mathematical probability. Standard probability axioms formalize probabilities in such a manner that – except in the extreme cases of 0 and of 1 values – if there is a probability from A to B, then there must also be a probability from B to A. That this might impose an unreasonable requirement on those conceptions of probability that incorporate causation, however, becomes apparent from the consideration that, when A is the cause of B, it is highly unusual, not to say impossible, for B to be the cause of A. So it may be necessary to qualify this condition for causal propensity conceptions.

In addition to the requirement of admissibility, Eells endorses the condition that an adequate explication must be conceptually adequate. This sounds like the merge of Hempel's conditions of semantic relevance and pragmatic significance, but Eells goes farther by endorsing a requirement of independent understanding that falls beyond Hempel's requirements. Eells distinguishes between two different species of accounts of both kinds, encompassing "actual" frequency and "hypothetical" frequency concepts as well as "long run" and "single case" conceptions of causal propensities. Eells argues that, while all of these views can be formulated in such a way that they satisfy formal conditions of admissibility and syntactical determinacy, none of them can successfully satisfy both the conditions of conceptual adequacy and of independent understanding, where to the extent to which they satisfy one of these conditions, they fail to satisfy the other. He draws parallels with developments in modal logic and possible-worlds semantics to suggest that frequency interpretations as species of correlations cannot satisfy the conditions for qualifying as candidates for natural laws, while propensity interpretations, although plausible candidates for natural laws, are unable to satisfy the conditions of independent understanding in relation to the semantics they imply.

Part I: Alternative Conceptions of Probability

Probabilistic Causality and Causal Generalizations

Daniel M. Hausman

Other important questions are raised by Daniel Hausman in the context of some of the special sciences, such as economics. He suggests that the truth of causal generalizations, such as "Smoking causes lung cancer", are independent of (what he calls) "metaphysical questions" concerning the nature of probabilistic causal relations like those described in quantum physics. Causation is a three-place relationship between a cause C, an effect, E, and a set of causally homogeneous background conditions, K. Even if this relationship were deterministic in the case of smoking and lung cancer, the relevance of smoking to cancer would still depend on the background conditions, which may in turn differ from person to person. In some people smoking could be a deterministic cause of lung cancer, while it is irrelevant or even a preventative in others. Whether the causal relations between smoking and lung cancer are indeterministic, in his view, is irrelevant. Not knowing the relevant factors in the background conditions or the facts about individual, the generalization, "Smoking causes lung cancer", expresses an average tendency.

There are various kinds of "averages", of course, including means, modes, and medians, where a football team could average 200 lb per player when, in fact, no member of the team weighs 200 lb. In some contexts this may matter more than in others. He also draws various distinctions between "relevance" and "roles", where in relation to a person's death, a poison and its antidote may both be relevant in making a difference to their death, while their roles are opposite, since one promotes death and the other inhibits it. Practical causal generalizations are, of course, concerned with causal role. He also discusses the difference between "variables" and "values of those variables", "homogeneous" and "heterogeneous" circumstances, and the "type/token" distinction, which he takes to be especially important in this context. His attention is mainly focused on (what he calls) "practical causal generalizations", which may include *ceteris paribus* clauses, typified by the claim, "(Wearing) seatbelts saves lives", which is usually true but sometimes not.

Hausman criticizes the views of Patrick Suppes, Nancy Cartwright, Paul Humphreys, Ellery Eells, and John Dupré, on the ground that they conflate the problem of specifying what it is for C to be a cause of E within some causally homogeneous background with the different problem of explicating causal generalizations concerning heterogeneous circumstances. He criticizes Suppes, Cartwright, Humphreys, and Eells, on the one hand, for assuming that practical causal generalizations make claims about indeterministic causation, and he criticizes Dupré, on the other, for making the opposite mistake of supposing that probabilistic causation is just a matter of averaging effects across heterogeneous background circumstances. Statistical generalizations across conditions that are not homogeneous could reflect either the operation of multiple but different deterministic causes or the operation of multiple indeterministic causes – or some combination both of deterministic and of indeterministic causes. But his analysis is meant to separate distinct problems rather than to resolve them.

The Possibility of Infinitesimal Chances

Martin Barrett

Among the more subtle but fascinating questions that arise within this context is the possible distinction between impossibilities and improbabilities of the value of zero. There are different modes of modality, of course, including those of logical possibility/impossibility/necessity, physical possibility/impossibility/necessity, and historical possibility/impossibility/necessity, where events satisfy corresponding conditions when their occurrence is consistent with/inconsistent with/required by the laws of logic, the laws of nature, and the history of the world up until some particular time, respectively. In the discussion of alternative interpretations of objective probability, where hypothetical frequencies and causal propensities are under consideration, the idea of identifying "chance 0" with physical impossibility and "chance 1" with physical necessity appears to be appealing. As Martin Barrett explains, formal difficulties arise when we combine this idea with a picture of a symmetrical space including an infinite number of outcomes and with standard axiomatizations of probability. Which means that interpretations that incorporate definitions of this kind are not even admissible.

One type of modern approach to these difficulties involves the notion of infinitesimal probabilities and nonstandard analysis in mathematics. Barrett describes the issues, attempted "infinitesimal/non-standard analysis" resolutions and technical difficulties that seem to have been overlooked by proponents of such resolutions. The approach follows an idea of David Lewis, who suggests that "zero chance" is no chance and that nothing with zero chance ever happens, identifying zero chances with physical impossibilities. He then treats "infinitesimal chance" as some chance. The treatment of infinitesimals in this fashion, Barrett observes, would potentially be welcomed by both classical and Bayesian statisticians, because classical statisticians cannot form likelihood ratios with zero probabilities and Bayesian statisticians cannot condition on learned propositions of probability zero. So there are positive reasons that motivate exploring whether or not using infinitesimal probabilities can produce a suitable probabilistic mathematics.

Application of these modern mathematical ideas to the problem in question, it turns out, is not as straightforward as one assume. Barrett argues that, however attractive the idea may appear, there are no suitable regular probability distributions with infinitesimals. On perfectly general grounds, he offers a proof that a suitable probabilistic mathematics requires differentiating zero probabilities from impossibilities. However plausible it might appear to have every value for the probability of a possible event in the [0,1] interval a positive, if infinitesimal, value, at least one such value in [0,1] must be zero. While Barrett's paper is highly technical mathematically, many parts are "reader-friendly" and accessible, with important implications for understanding probability in several contexts, including its role within decision theory in particular.

Probabilistic Metaphysics

James H. Fetzer

The criteria of adequacy that Salmon (1967) proposed were: (a) admissibility, where relations characteristic of mathematical probabilities (principles of summation, and multiplication, for example) must be satisfied; (b) ascertainability, where it must be possible to subject probability hypotheses to empirical evaluations using statistical tests; and (c) applicability, where the values of probabilities must be predictively significant for the long run, the short run, and the single case. James Fetzer argues that the single case propensity conception, according to which physical probabilities are probabilistic dispositions to produce (or "bring about") specific outcomes on single trials, can satisfy these criteria, where "long runs" and "short runs" are envisioned as infinite and finite sequences of single cases, but not conversely. Assuming they are equal and independent, this conception justifies classic statistical theorems, such as the Bernoulli and the central limit theorem, relating propensities and frequencies.

Thus, relative frequencies with specific values are the expectable outcomes of single-case propensities with similar values when they are subject to numerous repetitions. The values of these causal tendencies are the objective properties that can explain the occurrence of corresponding relative frequencies. And observed relative frequencies can provide suitable empirical evidence for statistical tests of propensity hypotheses, especially within the framework of "inference to the best explanation", to which we shall return. In abbreviated form, therefore: propensities can predict frequencies; propensities also explain them; and frequencies are evidence for propensities. The single-case conception, unlike any short or long run alternative, applies no matter whether the world's history is short or is long. And, unlike hypothetical frequency interpretations, it provides an ontological justification for specific distributions over hypothetical extensions of the world's history by subjunctives and counterfactuals.

In collaboration with Donald Nute, Fetzer developed a possible worlds semantics to capture the formal properties relating propensities and frequencies (Fetzer and Nute 1979, 1980). Based upon the conception of permanent properties and dispositions, they elaborated a semantics that employed infinite sequences of infinite sequences in order to illustrate that single case propensities no longer guarantee that long run frequencies must occur with limits that equal their generating propensities, but that such outcomes are overwhelmingly likely for the vast majority of infinite sequences on the basis of the kinds of normal distributions that are statistically to be expected. Indeed, this semantics invited Eells' objection that, without some principled method for ordering the members of the infinite set of infinite sequences, any distribution could be produced. Fetzer's response, in principle, is that we do not have access to any such infinite sequences in any case, where what we mean by propensity hypotheses has to be established by stipulations that distinguish between them, as he does here. The semantics thus illustrates the expectable consequences across infinite sequences of infinite sequences.

Part II: The Objectivity of Macro-probabilities

Chance and Necessity: From Humean Supervenience to Humean Projection

Wolfgang Spohn

The paradigm endorsed by Wolfgang Spohn affirms the existence of deterministic laws, where a state of affairs is physically necessary if and only if its occurrence is completely determined by its past. These laws generalize over their instances so that a state of affairs is also physically necessary if and only if its occurrence is entailed by the laws and its past. A state of affairs is chancy, by contrast, when its occurrence is only partially determined by its past "to some degree". Objective probabilities may be appropriately envisioned as single-case propensities, but of a special kind, namely: as "propensities of the entire world as it has developed up to now to realize not only this or that current state of affairs, but in effect this or that entire future evolution." Although this approach may exercise a certain attraction, it appears to disregard the evident consideration that only some, but not all, of the features of the world make a difference to specific outcomes, which means that most of the features of the past do not exercise any causal influence upon present events. Indeed, propensities satisfy the Markov condition where present causes bring about present effects (Fetzer 1983).

Spohn is not alone in advancing this conception, however, which is also found in the work of David Lewis, among others. He endorses Lewis' Principal Principle, which holds that, when the chance for an outcome happens to be known, then our degree of belief in the occurrence of that outcome should have the same value. He thus relates objective probabilities as single-case propensities to subjective probabilities as degrees of belief. Relative to Schema S above, this means that, if a probabilistic law of the form, "P $(B/A) = p$", happens to be known, then the probability of the logical link between the explanans and the explanandum, [p], should have the same value. Indeed, when the value of [p] equals the value of p, under the same complete sets of initial conditions, it is appropriate to envision [p] as a logical probability, which both denotes the relative frequency with which explanandum events of that kind will tend to occur but also denotes the degree of entailment with which that explanandum may be said to "follow from" that explanans. When there is a deviation between them, then the value of [p] represents a subjective probability, where the distance between them measures the extent to which subjective beliefs depart from objective expectations.

He raises many interesting questions and makes many valuable points in discussing the nature of Humean supervenience, which establishes an ontological foundation for drawing inferences about the future on the basis of knowledge of the past. His denial of the criterion of unification as the basis for identifying laws of nature – for which the laws of nature correspond to the theorems that are derivable from the simplest and strongest systematization – seems to be especially well-taken. He also denies Humean supervenience, maintaining that, even given complete knowledge of particular facts, it is "actually unfeasible to precisely detect chances". Without attempting to summarize the subtle and complex issues that he addresses, it may be worth noting that the kind of minimal-change possible-world semantics that Lewis and others endorse may not be the appropriate kind of semantics for scientific conditionals as singular sentences and generalizations attributing subjunctive and counterfactual properties and relations to objects and events in the world. When the antecedents of nomological conditionals are required to satisfy a requirement including the presence and the absence of complete sets of relevant conditions, in relation to which the propensities of their outcomes are constant, maximal-change semantics may be preferable (Fetzer and Nute 1979, 1980).

Evolutionary Theory and the Reality of Macro-Probabilities

Elliott Sober

Elliott Sober offers an extensive exploration of the philosophical consequences of the thesis of "mereological supervenience", according to which the precise properties that all micro-particles have at any specific time uniquely determine the properties that all macro-objects have at that same time. It is a form of part/whole supervenience, for which "micro determines macro". With respect to the converse relationship in which macro determines micro, Sober assumes that specific macro-states are often multiply-realizable at the micro-level. Thus, the temperature of a gas might be brought about by different distributions of its kinetic energy to its constituent molecules. Another would be that the same beliefs might be stored as mental dispositions in different fashions in different brains. Sober gives the problem a "probabilistic twist" by translating it into an analysis of the relationship between micro-probabilities and macro-probabilities.

If an event has the same probability, regardless of whether we conditionalize on the macro-state or the micro-state of the system, where their probabilities agree, then it is permissible to use the macro-probabilities for rendering predictions. If the macro- and micro-probabilities disagree, however, the micro-probabilities ought to be preferred if the micro-description entails the macro-description. Both principles may be regarded as based upon the epistemic requirement of total evidence. However, these principles might also be construed as jointly implying that micro-probabilities have an ontic status that macro-probabilities cannot equal, where, when they have divergent values, the macro-probabilities should not be supposed to describe "objective matters of fact". Sober takes his purpose to be to demonstrate that this inference is not correct and that a more adequate ontology vindicates the objectivity of macro-probabilities.

There are interesting questions about ontology that arise if "macro-probabilities" are interpreted as frequencies and "micro-probabilities" as propensities, but Sober does not pursue them here. He argues that indeterminism "percolates up", meaning that indeterminism at the micro-level entails indeterminism at the macro-level. He also maintains that, when the macro-probabilities are fixed, they almost never "screen off" the micro-probabilities, unless those macro-probabilities happen to be deterministic. This outcome, Sober contends, undermines an influential argument for reductionism advanced by Hilary Putnam. The kinds of cases that concern him the most, however, arise within the philosophy of biology, where he argues against the thesis that an organism's phenotype "screens off" its genotype from the organism's survival and reproductive success. An important dimension of his study is his discussion of differences between explanations and predictions in relation to macro-probabilities that matter in evolutionary contexts.

Is Evolution an Optimizing Process?

James H. Fetzer

While Sober emphasizes that evolutionary theory is "awash with probabilities", he does not commit himself here to one or another interpretation of probability as an objective property of the world. Fetzer, by comparison, advances an analysis of the hypothesis that natural selection qualifies as an optimizing process, which, he contends, cannot be sustained on the basis of either frequency or propensity conceptions. Thus, optimizing processes characterize systems that invariably select solutions to problems that are at least as good as any alternative solution, while satisficing processes characterize those that select solutions that are "good enough" but which may have alternatives that are even better. Since the values or utilities that matter within the context of evolution are those of survival and reproduction, one organism may have "higher fitness" than another when it has a higher probability of offspring reproduction than does another. Since nature cannot be expected to select solutions to problems that are not available, the question at stake should be in relation to existing gene pools and environments.

Whether or not optimizing may be realized in nature becomes much easier to assess when the conditions that it requires are made explicit. These conditions depend upon gene expression and the mechanisms of genetic change in populations. Of particular importance are the rate at which selection can alter the genetic structures, the amount of additive genetic variance present at the start, gene flow, the rate at which conditions change, and the random effects that affect the process, including genetic drift. Most optimality models presume sexual reproduction and the absence of pleiotropic effects, where changes in single genes can affect multiple traits. More importantly, however, the emergence of optimal traits requires infinite time and infinite populations. Which means optimality models provide idealized conceptions of what might happen "in the long run" if these conditions were ever satisfied rather than a descriptive explanatory framework for understanding evolution. A satisficing model, by contrast, provides a foundation for envisioning the emergence of optimal adaptations as a potential "long run" product of a "short run" process operating across finite populations and times.

The benefits of this exchange of paradigms would appear to be profound. If optimal adaptations only emerge as special limiting cases under idealized conditions, then it should be apparent that the products of evolution that emerge during any merely finite segment of the history of the world are never optimal, unless they happen to have appeared by virtue of fortuitous conditions that may have occurred by chance or by design. If this is indeed the case, then satisficing appears to be the strongest conception that generally applies. Fetzer argues further that envisioning optimizing in terms of hypothetical long-run frequencies does not salvage the situation, where the most promising approach to understanding probabilistic evolutionary phenomena arises from adopting the single-case propensity conception, which applies no matter whether the world's history is short or is long. Correctly understanding fitness as a propensity, however, requires an appreciation that "higher fitness" as a propensity does not guarantee a positive correlation between fitness and reproductive success. It also affords a framework for understanding the role of causal propensities as basic features of the world's structure that contributes to objectivity and realism in science.

Part III: Probabilities as Explanatory Properties

Propensity Trajectories, Preemption, and the Identity of Events

Ellery Eells

Eells elaborates and defends a propensity theory of singular causation based upon five principles. First, it is a theory of singular causation, by which he means that it applies to specific individual cases (say, "Harry's smoking caused his heart attack" as tokens or instances of type generalizations, such as "Smoking causes heart attacks"). Second, events are understood very broadly as exemplifications or instantiations of properties at specific places and times (which will be formalized as "Xx", where "x" stands for a place/time tuple; alternatively, they could be formalized as "Xxt", where "x" stands for a thing and "t" for a time). Third, this is a probability-increase theory, but not in the ususal type-level sense, where only factors that bring about increases in the probabilities of outcomes qualify as causes. Instead, this account focuses upon the actual evolution of the probability of specific token effects from the beginning of the instantiation of their specific token causes. A probability trajectory, therefore, is the shape of the time/probability graph that represents these changes across time.

The conception of probability that he embraces here makes probabilities relative to populations of different kinds. So the incidence of heart attacks among humans will differ from its incidence among laboratory smoking machines. Fourth, he offers two qualifications, where one is intended to cope with "spurious correlations" in which the apparent relations between events may be brought about by a common cause; and the other is intended to define "positive", "negative", and "neutral" causal factors in relation all causal background contexts and where those that occur in some but not in others are called "causally mixed". Fifth, he introduces terminology and constraints to capture the token-level evolution of trajectory values. He distinguishes between four kinds of causal significance, which he labels "because of", "despite", "independently of", and "autonomously of", respectively. This apparatus thus allows him to explore multiple applications of this conception of probability to issues of causation and explanation.

One kind of problem Eells discusses is that of causal preemption, where there may be more than one apparent "explanation" for an outcome and it becomes important to differentiate between them. If two persons want to kill another as he sets forth to cross a desert, one of them might poison his water, while the other punctures his canteen. Although the poisoned water drains out, the lack of water causes his death. He demonstrates that the propensity trajectory theory offers the right solution, since it establishes that his death by dehydration was causally independent of the poison in his canteen. Poisoning the water thus turns out to be causally irrelevant to this death. If the man had drunk the poison before his canteen ran dry, of course, then the poison would have made a difference and might have caused his death. Thus, in this case, the cause of death was by dehydration despite the canteen having been poisoned. And he demonstrates that similar appropriate results occur from the application of this theory not only to cases of preemption but also of symmetrically overdetermining events. His study thus provides strong conceptual support for the propensity trajectory conception.

Miraculous Consilience of Quantum Mechanics

Malcolm R. Forster

Malcolm Forster focuses upon Occam's Razor as a rule of thumb that maintains that, in the search for truth, "Entities are not to be multiplied beyond necessity". The meaning of this principle has invited alternative interpretations, where Forster raises several interesting questions, such as "Do probabilities qualify as entities?"

and, if so, “Should probabilities not be multiplied beyond necessity?” He applies Occam’s Razor to probabilities relative to double-slit experiments in quantum mechanics, which leads to false predictions Yet he discovers that quantum mechanics appears to be parsimonious with respect to other kinds of entities. He also discusses the Bell theorem, which has been widely assumed to establish there are no “hidden” variables that affect outcomes in quantum mechanics that would turn indeterministic phenomena into deterministic phenomena and has led some, such as Bas van Fraassen, to argue that there are no causal explanations for at least some quantum phenomena.

A student of William Whewell, Forster elaborates upon his consilience of inductions, emphasizing how “good theories” not only relate themselves to the data in a narrow domain but also “tie together” results in other domains. This occurs when either (a) the theory accommodates one kind of data and predicts data of yet another kind or (b) the theory accommodates data of two different kinds but yields results that reveal constants or laws that tie them together. His primary objective in this paper is to demonstrate that quantum mechanics can be viewed as achieving a consilience of the kind Whewell had in mind. This, he argues, is a suitable “starting point” for the development of a realist interpretation that is “more parsimonious” than any hidden variable interpretation of the same phenomena. And he suggests that the consilience of inductions provides a criterion for the applicability or not of causal explanations, which leads to the conclusion that causal explanations can be “ruled out” for at least some quantum mechanical phenomena. But of course it depends upon the meaning of the phrase, “causal explanation”, where there may be other kinds of *probabilistic* causal explanation that do apply to these quantum phenomena.

Those who adopt the view that “God does not play with dice!”, as Einstein reportedly observed, are inclined to construe probabilities in quantum mechanics simply as “measures of ignorance”, where future discoveries may well reveal the deterministic mechanisms that underlie our probabilistic descriptions. Bohr, by contrast, thought that the indeterminacies reflected by the laws of radioactive decay, for example, are permanent features and are not going to be replaced, no matter how much more we learn about the world. In discussing the double-slit experiment and the uncertainty principle, Forster contends that modeling in quantum mechanics is very different from curve fitting because it relies upon operators rather than variables. Possibly hidden variable theorists have failed to show that quantum probabilities are measures of ignorance, not because there are no “hidden factors”, but because those factors cannot be represented as *variables*. He also contends that a new geometrical interpretation of quantum mechanics better explains the use of operators in quantum theory and may eventually succeed in providing a realist interpretation of quantum phenomena, where probabilities in quantum mechanics really are simply measures of our ignorance.

Probability and Objectivity in Deterministic and Indeterministic Situations

James H. Fetzer

It appears to be rather difficult to defend a realist interpretation of quantum probabilities in the absence of an exploration of the alternative possibilities for fulfilling that role. Fetzer contends that, among the three most prominent interpretations of probability – the frequency, the subjective, and the propensity – only the third accommodates the possibility of providing a realist interpretation of ontic indeterminism. He discusses the criterion of reality introduced by Einstein, Podolsky, Rosen, according to which, if, without disturbing a system, we can predict with certainty the value of some physical quantity, then there is an element of physical reality corresponding to that quantity. And he explains why this formulation appears to be flawed by covert commitments to determinism, according to which, if we can predict every attribute of a system S for every moment t with certainty, then (a) our description of S is complete and (b) S is governed by strict causal laws. If determinism is built into a criterion of reality, after all, indeterministic phenomena are not going to qualify, no matter how real.

He offers an alternative formulation that creates the opportunity for at least some of the phenomena of physics to be both indeterministic and real, by asking whether, if, without in any way disturbing a system S, we can predict, not with certainty but with some probability less than 1, the value of a physical quantity, can there not still exist some element of physical reality corresponding to that physical quantity? Einstein's criterion as a sufficient condition for the existence of an element of reality, after all, should not be confounded with, say, a sufficient definition as the weakest criterion of reality, which would pose both a necessary and a sufficient condition for existence as an element of reality. For ontic indeterminism to be theoretically possible, it must not be logically inconsistent to suppose (a') that our description of S is complete even though (b') the system S is not governed by "strict causal laws". Indeed, it is crucial to realize that the laws involved here can be causal even though they are not strict, because they reflect irreducibly probabilistic causation, not epistemic indeterminism.

Fetzer argues that, across a broad range of quantum phenomena, only one of the three approaches can accommodate both explanatory and predictive kinds of indeterminism. The frequency criterion of statistical relevance, for example, demands that properties that make a difference to the frequency with which specific outcomes occur have to be taken into account; but since every event happens to be unique, rigorous application of that criterion produces only degenerate probabilities of zero or of one for any outcome. It therefore cannot accommodate either predictive or explanatory indeterminism. The subjective criterion of evidential relevance, by contrast, can accommodate predictive indeterminism by virtue of an absence of knowledge, but once the occurrence or non-occurrence of an event becomes known, its subjective probability necessarily becomes equal to zero or to one. Hence, it can accommodate predictive indeterminism but not explanatory indeterminism. Only the propensity interpretation permits probabilistic hypotheses assigning non-degenerate probabilities to known or unknown outcomes to be true. The propensity approach thus qualifies as the only account that provides the foundation for a realistic interpretation of fundamental quantum phenomena.

Part IV: Probabilities in Inference and Decision

How Bayesian Confirmation Theory Handles the Paradox of the Ravens

Branden Fitelson and James Hawthorne

In addition to its role in philosophical endeavor to understand the "chanciness" of the chancy character of the world, the concept of probability has also been important in philosophy in relation to the endeavor to understand the nature of rational inference and rational decision. The theory of rational inference, for example, encompasses the phenomena of "theory choice" and of "hypothesis confirmation" in the philosophy of science, while the theory of rational decision also encompasses choice between laws and theories, especially principles widely assumed required to understand economic, social, and individual behavior. In the logic of hypothesis confirmation, two general kinds of question arise: (1) Which of two different pieces of evidence, call them "e" and "e'", confirms a given hypothesis h more than the other, and why?; and (2) Which of two hypotheses, h and h', does given evidence e confirm more than the other, and why? A famous problem related to the first kind of question is Hempel's "paradoxes of confirmation", also known as the "raven paradox" and now as "Hempel's paradox". A famous example of the second kind of problem is Nelson Goodman's puzzle about the predicates "grue" and "bleen". puzzle. Fitelson and Hawthorne attack the first.

Fitelson and Hawthorne advance a new "Bayesian" approach employing subjective probabilities to resolve the paradox. A natural idea for the theory of confirmation of hypotheses is that a statement or hypothesis of the form, "All Fs are Gs", is confirmed by the observation of things that are both F and G. Another natural idea is that, when evidence e confirms an hypothesis h and that hypothesis h is logically equivalent to another hypothesis, h', then e also confirms h'. Hempel noticed that the hypothesis, "All ravens are black", in standard extensional logic, turns out to be equivalent to the hypothesis, "All nonblack things are nonravens". The original formulation of what is presumably the same hypothesis should therefore be confirmed by the observation of a black raven, while the second would be confirmed by the oberation of a nonblack non-raven, such as the space between this and the next sentence of this book. But if logically equivalent hypotheses are confirmed by the same evidence, then observing the space between these sentences should not only confirm the hypothesis that non-black things are non-ravens but also its logical equivalent that all ravens are black.

Fitelson and Hawthorne pursue the application of probabilistic, Bayesian confirmation theory to this paradox, motivated by the belief that, while observations of nonblack nonravens confirm the hypothesis that all ravens are black, the degree to which they do so is minuscule in comparison to the degree to which observations of black ravens confirm the hypothesis. While others have also taken this path, they advance a novel variation that appears to be more promising. Fitelson and Hawthorne provide a very meticulous description of the paradox, including the assumptions, the arguments, and the conclusions involved. While all Bayesian/probabilistic approaches to the problem identify specific sufficient conditions on probability functions that support intuitively correct answers, the Fitelson–Hawthorne approach is to "zero-in" upon less restrictive sufficient conditions, which makes their position both technically and philosophically interesting, while offering a real advance over other proposed probabilistic solutions.

Learning to Network

Brian Skyrms and Robin Pemantle

The theory of choice includes scientific theory choice and the theory of individual decision in ordinary contexts in general, but also the choice of strategy when there are multiple agents or decision makers, who may be cooperating or even competing. In the theory of choice, the focus is upon rationality of action as a relation between a person's behavior and their beliefs, desires, and preferences. It is prescriptive rather than descriptive. In a formal theory of rational choice, beliefs are typically modeled by subjective probability assignments, desires by numerical utility functions (which attach to potential outcomes of possible actions dependent upon states of the world). Preferences between alternative outcomes and the expected utilities of actions that might be adopted are supposed to arise as a function of personal probabilities and subjective utilities. In what are called "strategic" or "game theoretic" contexts with multiple agents or "players", the participants may also anticipate one other's moves.

An important concept is that of the "Nash equilibrium", where no player can do any better by switching their choice as long as none of the others do so. However, the tools of formal game theory have application outside the realm in which deliberate choices are made on the basis of conscious consideration of probabilities and utilities. In their paper, for example, Brian Skyrms and Robin Pemantle investigate, from a probabilistic, game theoretic point of view, the explanation and evolution of actual behavior patterns of individuals (or of how one individual behaves towards others) within a species (or more generally, within any group of individuals) on the basis of what they take to be two basic factors, namely: (i) the history of outcomes of specific interactions with other individuals with whom given individuals have interacted in the past; and (ii) the strategies adopted by the individuals under consideration, where (i) and (ii) can change as time goes by. Of course, the history of interactions among individuals will be different as time goes by, but the strategies employed can change, for example in response to previous interactions. The situation is therefore dynamic.

Skyrms and Pemantle employ the tools of formal game theory in their investigation. Three of the most salient features of their investigation are, first, their approach is probabilistic rather than deterministic; second, their approach does not require the adoption of personal probabilities as subjective "degrees of belief" but is developed on the basis of "probabilities" and "utilities" as a game-theoretical framework for explaining the evolution of patterns of behavior; and, third, their mathematical application of the idea of dynamics in their analysis includes the use of computer simulations of histories of interaction and histories of strategy choice with the aim of investigating how these two factors interact. The result is a broad framework for the study of the evolution of group and even species behavior, which is normative in character but also explanatory when the assumptions on which it is based happen to be satisfied by the members of those groups and the populations of those species.

Probabilities in Decision Rules

Paul Weirich

In "classic" Bayesian decision theory and subjective probability theory, persons who are rational are assumed to possess "degrees of belief" that satisfy some standard axiomatization of probability. (Degrees of belief that accommodate this condition are often referred to as "personal probabilities" rather than merely "subjective".) This is clearly an idealization, however, which is seldom completely satisfied. One standard axiom of probability, for example, maintains that the probability of any logical truth is equal to unity (when "probability" is understood as taking propositions as objects to which it may be assigned). Interpreted subjectively as "degrees of belief", this means that any agent who qualifies as "rational" would have to have the degree of belief of unity in any proposition that qualifies as a logical truth, no matter how complicated the proposition or the language in which it is expressed. That would encompass an infinite number of propositions of arbitrary complexity, a requirement that humans with finite capacities cannot be expected to satisfy. It represents an idealization.

Another standard axiom (called "additivity") is that the probability of any disjunction of the form, "p or q", is equal to the sum of the probabilities of those disjuncts, "p" and "q", when p and q are logically incompatible. Interpreting "probability" subjectively, this would seem to imply that rational agents must be sensitive to all possible logical incompatibilities in order for their subjective probabilities to satisfy this requirement. Again, of course, this is not a realistic expectation for any human mind, yet we would not want to consider most humans as "irrational" on that account alone. Paul Weirich suggests that a promising way to make the theory of subjective probability more realistic and evaluations of the rationality of decisions more fair or reasonable is to envision degrees of belief as attaching not to propositions, strictly speaking, but to each agent's way of "grasping" them. Since propositions are expressed by sentences, not only can two sentences be "grasped" as expressing the same proposition but, in different contexts, the same sentence can be used to convey different propositions.

No doubt, theories of rationality of belief as well as of rationality of action, like those of optimality in evolution, need to accommodate more realistic conceptions in which satisficing standards are generally applicable and maximizing the exception. Weirich investigates the idea of what might be called the "forgiveness" of even ideally rational agents to "grasp" the same proposition in different ways, and he suggests relativizing the rationality of decisions to an agent's assessments of probabilities of propositions in relation to the way in which they are actually grasped. He argues that this has the effect of generalizing the way in which the principles of subjective probability are to be understood, where the usual understanding is a special case. It certainly comes closer to reflecting the conditions of real human behavior. And he develops his own and others' thoughts about how to make rational decision theory and the theory of rational degrees of belief more realistic, less idealistic, for actual human contexts.

Epilogue

Propensities and Frequencies: Inference to the Best Explanation

James H. Fetzer

Fetzer elaborates an approach toward inference to the best explanation integrating a Popperian conception of natural laws together with a modified Hempelian account of explanation, one the one hand, and Hacking's law of likelihood (in its nomic guise), on the other, which provides a highly robust abductivist model of science that appears to overcome the obstacles encountered by its inductivist, deductivist, and hypothetico-deductivist alternatives. This philosophy of science clarifies and illuminates some fundamental aspects of ontology and epistemology, especially concerning relations between frequencies and propensities. Among the most important elements of this conception is the central role filled by degrees of nomic expectability in explanation, prediction and inference, for which this investigation provides a theoretical defense.

Thus, when scientific conditionals are understood as logically contingent subjunctive conditionals that attribute "permanent properties" to things that possess appropriate reference properties, both probabilistic and deterministic causal laws have the form of logically unrestricted generalizations. It is no longer a question of what percent of the reference class possesses the attribute of interest, but a matter of the strength of the causal tendency possessed by every member of that reference class. Permanent properties thus both explain and justify attributions of susperve-nience. Explanations differ from predictions insofar as explanations are adequate only when the properties they cite are restricted to those that are causally (more broadly, nomicly) relevant, but predictions can be adequate even when they are based upon causally (nomicly) irrelevant properties. While satisfying the requirement of maximal specificity (by specifying complete sets of relevant properties) is a necessary condition for the truth of lawlike sentences, the adequacy of explanations requires satisfying the condition of "strict" maximal specificity as well (by excluding explanatorily irrelevant properties).

These considerations have ramifications for the application of the Principal Principle. Knowledge of laws of nature, which cannot be violated and cannot be changed, takes predictive primacy over knowledge of relative frequencies that have obtained in the past. When we possess knowledge of single-case propensities, therefore, they ought to determine the values of corresponding degrees of belief for inference and decision. When knowledge of single-case propensities is unavailable, however, then degrees of belief should be determined by beliefs about corresponding relative frequencies. In cases where neither knowledge of single-case propensities nor knowledge of relative frequencies happens to be available, however, then decision making depends upon hypothetical reasoning or educated guesswork, where rationality of action tends to be decoupled from rationality of belief. Actions taken under conditions of this kind are not only extremely risky but are subject to the influence of psychology and ideology.

Significantly, the permanent property relation guarantees that, if a specific thing c of kind R has an attribute A as a permanent property by virtue of being a thing of kind R, then every thing x of kind R must have that same permanent property. And this implies that, in the case of permanent properties, inferences from specific instances to universal generalizations are valid (Fetzer 1981, p. 53). The empirical testability of lawlike hypotheses, moreover, establishes a foundation for resolving the problem of induction. With respect to the class of presumptive laws of nature, in particular, even if the world is as we believe it to be in these specific respects, there remains the logical possibility that it might change in the future. But if the world is as we believe it to be in these specific respects, it is not physically possible that it might change in the future. Which suggests, in turn, that certain classic philosophical problems about dispositions, probability, causation, and laws may have remained unresolved because methodological commitments to extensional languages and truth-functional logic have inhibited the adoption of more adequate but non-standard and intensional solutions.

Ellery Eells and James H. Fetzer

References

Fetzer JH, Nute D (1979) Syntax, semantics, and ontology: a probabilistic causal calculus. Synthese 40/3(March 1979):453–495

Fetzer JH, Nute D (1980) A probabilistic causal calculus: conflicting conceptions. Synthese 44/2(June 1980):241–246

Fetzer JH (1981) Scientific knowledge: causation, explanation, and corroboration. D. Reidel, Dordrecht/Holland

Fetzer JH (1983) Probabilistic explanations. In: Asquith P, Nickles T (eds) PSA 1982. Philosophy of Science Association, East Lansing, MI, pp 194–207

Hempel CG (1952) Fundamentals of concept formation in empiricial science. University of Chicago Press, Chicago, IL

Salmon WC (1967) Foundations of scientific inference. University of Pittsburgh Press, Pittsburgh, PA

Acknowledgments

The editors hereby gratefully acknowledge their gratitude to Springer Science and Business Media for permission to republish the following articles as chapters of this volume.

Fetzer JH (1983) Probability and objectivity in deterministic and indeterministic situations. Synthese 57/3(December 1983):367–386

Eells E (1983) Objective probability theory theory. Synthese 57/3(December 1983): 387–442

Fetzer JH (1988) Probabilistic metaphysics. In Fetzer JH (ed) Probability and causality. D. Reidel, Dordrecht/Holland, pp 109–132

Eells E (2002) Propensity trajectories, preemption, and the identity of events. Synthese 132/1–2 (August 2002):119–141

Fetzer JH (2002) Propensities and frequencies: inference to the best explanation. Synthese 132/1–2 (August 2002):27–61

Memorials

Ellery Eells (1953–2006)

Malcolm R. Forster

Professor of Philosophy, University of Wisconsin-Madison, until his untimely death on August 10, 2006, at the age of 52. He is survived by his wife, Joanne Tillinghast, son Justin, daughter Erika, father Thomas Eells, as well as three brothers and two sisters.

Born and raised in Los Angeles, Ellery completed all his education in California. First he went to Santa Barbara to study philosophy and mathematics, graduating as Outstanding Graduating Senior in Philosophy in 1975.

After that, he moved further north to the University of California, Berkeley, to earn a Ph.D. in philosophy in 1980. Except for a one-year visiting position at North Carolina State University, Ellery's entire working career was spent at the University of Wisconsin-Madison, from 1980 to the present.

Ellery first gained major recognition in philosophy from his book, *Rational Decision and Causality* (Cambridge University Press). The book was published in 1982 at the height of the uproar over Newcomb-style counterexamples to Bayesian decision theory. In it, Ellery developed the entirely novel argument that Bayesian decision theory can produce the same answers as the new causal decision theory so long as deliberation is viewed as a dynamical process.

Besides spear-heading this new line of research in decision theory, his work rekindled interest in old questions about the relationship between causality and probability. The paper he published with Elliott Sober in 1983, called "Probabilistic Causality and the Question of Transitivity" is still widely cited in this area.

Finally, this culminated in a major treatise called *Probabilistic Causality* in 1991. In the meantime, he was publishing numerous papers in confirmation theory; perhaps the best known is "Problems of Old Evidence", which first appeared in the *Pacific Philosophical Quarterly* in 1985. It has been reprinted twice since then.

Ellery won the American Philosophical Association's Franklin J. Matchette Prize for his book on probabilistic causality in 1995, after already receiving a John Simon Guggenheim Memorial Foundation Fellowship, an ACLS award, and numerous awards from the University of Wisconsin-Madison. More recently, he was elected to the Governing Board of the Philosophy of Science Association.

He has always been tireless in his service to the philosophy of science community, especially behind the scenes. For example, he normally wrote more than ten reviews and referee reports per year. His life was dedicated to philosophy.

To those who knew him personally, Ellery was a kind and gentle person, with a quiet but cheerful demeanor. Academically, had an unparalleled patience for details, which shows up very clearly in his published work. (He once told me that he had never had a paper rejected for publication!) His patience made him popular amongst the graduate students and colleagues who sought his expertise, and equally amongst those who were novices in his field.

Ellery was the person with whom you'd want to share committee work; he was hard-working and reliable, and would always have copious notes. When serving on the admissions committee, for instance, Ellery was always able to summarize the best points of every candidate. He was always looking for the good in everyone.

I feel privileged to have been his colleague, and his friend. A memorial session for Ellery Eells has been organized for the Pacific APA meetings, San Francisco, April 2007. The National Taipei University of Technology is opening a new research center called the Ellery Eells Memorial Center for Philosophy of Science and Professional Ethics.

University of Wisconsin-Madison Malcolm Forster

At the Memorial Gathering for Ellery Eells

Elliott Sober

We are here to mark a sad day, the death of Ellery Eells, by sharing with each other the memories we have of Ellery.

My name is "Elliott Sober." Ellery was my friend and he also was a wonderful philosophical colleague. We read each other's drafts of books and articles, talked about projects, and wrote some papers together. When I think of the philosophical work we did together, and of our time together just relaxing and enjoying each other's company – these are very happy memories.

Over our years working together, Ellery was a meticulous and penetrating critic of the papers I was working on. And he was *tireless*. Ellery would find a problem in an argument I was trying to develop, I would attempt to patch it up, and then he would look at the revision. Sometimes it was the same old problem, still there; at other times, my revision merely substituted a new problem for the old one that was there before. Ellery would point this out with great tact and patience, and the process would continue. When we academics get other academics to give us comments on our work, we have to be careful not to try their patience. One read of a paper is usually as much as we can expect. But Ellery was far more altruistic than this.

In 1987, Ellery and I travelled together to an international philosophy of science congress in Moscow. We got to Moscow by a circuitous route, stretching out the trip to visit some interesting places. We flew to Helsinki, then took an overnight boat across the Baltic to St. Petersburg, and then we flew to Moscow. We spent a week getting to Moscow and a week in Moscow, rooming together the whole time. After we got back to Madison, Ellery and I would enjoy reminding each other of little details from our trip. One of them involved an interpreter we had in Moscow; he had excellent English, though he consistently confused the expression "thank God" and "for God's sake." When the three of us got into a long line at a restaurant, the interpreter said, "We have a long wait ahead of us, thank God." And when we finally reached the front of the line, he said "We'll soon be having our meal, for God's sake." Ellery was an excellent travelling companion. His good humor, his noticing details, his pleasure at seeing new sights – all of these characteristics make

E. Sober (✉)
Gates of Heaven, James Madison Park Madison, Wisconsin September 27, 2006

me smile as I remember our beat-up hotel near Red Square, our visit to a monastery outside of Moscow, our walking around in the rain in a garden in St. Petersburg, and our seeing Lenin's body.

Ellery and I sailed together on Lake Mendota. Ellery learned to sail as a boy in his father's boat on Lake Arrowhead, near Los Angeles. I learned to sail after I started teaching here by taking lessons at Hoofer's, the student sailing club. When Ellery and I were out on a sailboat, I was often struck by the difference between learning to sail as a child and learning to sail as an adult. Ellery was relaxed while I was hyper-alert. Ellery took his time to do things and he did them efficiently; I would rush and bungle. Once we were out in the smallest boat that Hoofer's has, the tech dingy, and we were racing against another boat. Ellery and I lost because we changed course too often. I say "we," but maybe I should say "I," because Ellery knew better than to do this, but in his characteristically gentle way, he did not press the point. By his example, not his words, Ellery impressed on me the virtues in sailing of "less is more." Another time, Ellery and I went out sailing with Ellery's father. I recall standing on the dock and losing hold of the rope. But there was no reason to panic since Ellery's father was in the boat and easily sailed it back to the dock. Why get upset by such little things? Ellery did not, and he helped me learn not to do so.

Some of Ellery's former students have sent me emails containing their thoughts about Ellery. I'll read these to you after others here have said what is on their minds.

Comments from Some of Ellery's Former Students

1. Branden Fitelson (now teaching at University of California, Berkeley)

Ellery was a gentleman and a scholar. He was also one of my most cherished mentors. I took my first course with Ellery in 1991. It was on confirmation theory. I'm still working on that topic today, some 15 years on. That's no coincidence. He was pure gold. I often say to myself when preparing a lecture: "How would Ellery explain this? There must be an easier and simpler way." I'll never forget his uncanny ability to explain even the subtlest of concepts in the most simple and transparent ways. He made his students believe that they could understand anything (and with ease). Moreover, he never (not once in the time that I knew him) spoke in a disparaging way about the views of another philosopher. On the contrary, he spent almost all of his time in seminars reconstructing the views of others in the most charitable possible ways. Sometimes I wish he talked more about his own views. In this sense, his influence on his students and colleagues was profound, despite his unassuming and modest character. I think Alan Hajek summed things up beautifully in a recent email message. I hope Alan doesn't mind me including the following quotation here:

"When David Lewis died, I found some comfort in the fact that his work lives on. The same will be true of Ellery. He was a model in this respect, too, as you know. Again, his writings were right to the point – not showy or overblown, just insightful and crystal clear. My only criticism of him, if that's the right word, was that he seemed too modest to me. I wonder if he realized how good he was. One rarely encounters that kind of purity of intellect and character. It's something for us to aspire to."

I am organizing a memorial session for Ellery at the Pacific APA. I hope many people will be there to honor his memory. Rest in peace, Ellery.

2. Will Seaman (now an engineer at Hewlett Packard in Seattle)

As a former student, I feel deeply indebted and grateful to Ellery for his limitless patience and steady guidance through my graduate studies and dissertation. He was open to my rather too expansive ambitions, allowing me to explore topics that I think others might have discouraged, and he made no effort to push me in a direction or line of thinking that was more congenial to his own. Where he was firm and resolute was in his insistence that I not evade shortcomings in my reasoning and arguments, forcing me to rework and rethink. Those flaws that persist in the work I did under Ellery's kind supervision are truly my own, duly noted by Ellery and reflecting both his role as unrelenting advisor, but also his generosity and understanding of views that diverged from his own. I feel so very fortunate and privileged to have had Ellery as a teacher and advisor, and so terribly sad at the news of his death. I send my deepest sympathy to his wife and children, to his family, and to his many friends at this memorial gathering. We will all miss Ellery.

3. Mehmet Elgin (now teaching at Mula University in Turkey)

I am very sorry about Ellery's death and for the loss that this represents for his family, friends, students, and colleagues. Ellery was one of the smartest but also one of the most modest people I have ever known. It is very sad that philosophy of science has lost one of its most important figures. And it is doubly sad, since Ellery could have added to the important philosophical contributions he has already made some new and important contributions as well. As a last word, I can only say – "Good Bye, Ellery."

4. Kevin Brosnan (now teaching at University of California, Santa Cruz)

Ellery was one of the kindest, most gentle, and most generous people I have had the good fortune to know. When I asked him recently to be a member of my thesis committee, he agreed without hesitation, though he must have been feeling quite ill. Although under no obligation to do so, he read my dissertation carefully and posed detailed questions during my defense. Prior to this, he kindly agreed to attend a presentation of mine in the department. He came, having read my paper, and again raised insightful questions that helped me greatly. Over the years, no matter how simple-minded and confused my questions, Ellery always took the time to answer them thoroughly and thoughtfully, and without revealing how ill-conceived many of them probably were. Ellery's departure is a terrible loss to all of us. I will miss him greatly.

5. Sara Chant (now teaching at University of Missouri)

I am really sad about the loss of Ellery. I spent a great deal of time with him. He gave me a lot of encouragement and confidence about my work. And he was a really decent guy. He had a kind, gentle way about him, and his intellect was astounding. He was funny too though I never really could figure out what would make him laugh. He did laugh a great deal when I threatened to name my horse “Modus Pony”. It was fantastic. After that, he would always ask about the animals and he would laugh as he asked. He was a real sweetheart. I’ll miss Ellery a great deal.

Remembrances of Ellery Eells

Paul Weirich

Ellery and I entered the profession about the same time. We were both decision theorists engrossed by Newcomb's problem. Because we attended the same conferences, we quickly met and began corresponding about our work. Ellery sent me his dissertation. It was clear that his ideas would make a huge splash. What a great moment when Cambridge University Press published *Rational Decision and Causation*! It is a landmark publication in the vast literature on Newcomb's problem.

Ellery was incredibly good at fielding questions after his conference presentations. He constructed detailed and precise arguments for novel ideas on the spot. I was delighted when he volunteered to comment on an APA presentation of mine. He made scores of insightful points and turned his excellent commentary into an article for the *Australasian Journal of Philosophy*. Ellery was a steadfast source of ideas and encouragement. He stimulated many of my thoughts about decision instability and helped me progress in the profession.

When Ellery earned tenure at the University of Wisconsin, he built a new house there. Shortly after he and his family moved in, I remember talking with him on the phone and pausing from time to time so that he could move the sprinklers that were bringing newly sown grass seed to life. We shared stories about our families. His deep attachment to his wife and children were very evident. Our long conversation filled out my portrait of Ellery. It enriched my life to share that happy time with him.

I owe a lot to Ellery as a philosopher and a friend and miss him dearly. He was a prince.

Prologue

Objective Probability Theory Theory*

Ellery Eells

Conditions of Adequacy

Philosophical discussions on the topic of probability have mainly focused on two kinds of issues, the first having to do with the *concept* of probability and the second having to do with *methodological* standards which an interpretation of probability (i.e., a philosophical theory about the nature of probability) itself must satisfy if it is to be an adequate interpretation. As to the first kind of issue, philosophical theories of probability must endorse some more or less vague intuitions about what kind of thing probability is, and the conception of probability offered must accommodate the intuitions endorsed. For example, it's generally thought that probability must satisfy some sort of standard axiomatization, such as Kolmogorov's; it's often thought that physical probability must be objective in that probability values are correct or incorrect independently of anyone's state of knowledge or beliefs about the correctness of the values; on the other hand, it's also often thought that probability can only be a measure of our ignorance; it's generally thought that probability must have predictive significance and appropriately reflect certain causal features of the physical world; and it's generally thought that probability, whatever it is, must be applicable to the "the single case", in particular, in contexts of rational decision making and of probabilistic explanation.

Any theory of probability which doesn't endorse or accommodate sufficiently many such intuitions wouldn't constitute an interpretation of *probability*, but rather of something else if anything else. Much recent philosophical work on probability has been devoted to developing conceptions of probability that are sensitive to certain intuitions and to arguing that one or another proposal does *not* adequately accommodate some such intuitions. Now this is not to deny, of course, that there may be several different useful interpretations of probability. And I don't mean to

E. Eells (✉)
Department of Philosophy, University of Wisconsin-Madison, Madison, WI 53706
e-mail: jfetzer@d.umn.edu

*This paper was written, in part, under a grant from the Graduate School of the University of Wisconsin-Madison, which I gratefully acknowledge. I would also like to thank James H. Fetzer for many useful suggestions which improved an earlier draft, and Michael Byrd for useful discussions.

E. Eells and J.H. Fetzer (eds.), *The Place of Probability in Science*, Boston Studies in the Philosophy of Science 284, DOI 10.1007/978-90-481-3615-5_1,

assert that all of the above intuitions are relevant to every conception of probability. Rather, the point is just that at least one desideratum relevant to assessing the adequacy of a philosophical interpretation of probability is that the *concept* offered must be theoretically adequate in some appropriate sense.

I shall divide the conditions which an interpretation of probability must satisfy in order for it to be theoretically adequate into two parts. I shall call the condition that the interpretation of probability offered must satisfy some standard axiomatization the *condition of admissibility*. This follows the terminology of Salmon (1967, pp. 62–63), except that Salmon uses the term 'criterion' instead of 'condition'. The condition that the concept offered must be otherwise theoretically adequate I shall call the *condition of conceptual adequacy*. This condition roughly corresponds to Salmon's "criterion of applicability", the force of which, he points out, may be succinctly summarized by Bishop Butler's aphorism, "Probability is the very guide of life".

The second kind of issue in philosophical discussions of probability has to do with philosophical methodology and general standards, or conditions of adequacy, which a philosophical theory of probability must satisfy independently of the particular conception of probability proposed. Thus Suppes (1973, 1974) has recently criticized Popper's (1959) propensity interpretation on the grounds that it does not *formally characterize* probability in terms of ideas *understood independently* of quantitative probability, supposing that any adequate interpretation of probability must do this regardless of the particular conception of probability offered, whether it be subjective, Bayesian, single case propensity, hypothetical limiting frequency, etc. And Suppes (1973), Kyburg (1974, 1978), and Giere (1976) have recently attempted to develop the propensity interpretation in such a way that it satisfies what Giere (1976) calls "Suppes' Demand for a Formal Characterization of Propensities".

In the second section of this paper, I will elaborate Suppes' demand, dividing it into two parts. The *condition of formal adequacy* will demand that any adequate interpretation of probability provide us with a definition of its characteristic kind of axiomatized structure $\mathscr{C}$ – one instance, $\mathscr{S}$, of which will be the interpretation's "intended model", as explained below – where certain features of such a structure must be "representable" in terms of a probability function P on an appropriate structure $\mathscr{B}$, generally a part of a characteristic structure $\mathscr{C}$, as explained below. It will be seen that satisfaction of this condition is importantly related to the testability of a theory of probability in connection with satisfaction of the condition of admissibility. The *condition of interpretation/idealization* will take seriously the part of Suppes' demand – not adequately appreciated, I think, in Suppes (1973), Kyburg (1974, 1978), and Giere (1976) – that probability be characterized in terms of things *understood independently* of quantitative probability. This part of Suppes' demand itself has two parts: first, that the things in terms of which probability is explicated be *understood*, and second, that they be understood *independently* of the concept of quantitative probability. The condition of interpretation/idealization will demand specification of the intended model $\mathscr{S}$, alluded to above and explained more fully below, and that $\mathscr{S}$ be a model of, an *idealization* of, some understood concepts, objects or phenomena – ideally, features of the observable, empirically accessible world – so that those things constitute an *interpretation* of the constituents of $\mathscr{S}$ which at least roughly obey the relevant axioms which $\mathscr{S}$ obeys. The "understood" part of the second part of Suppes' demand will be satisfied if we "understand" that those things obey the relevant axioms, and the "independently" part of the second part of Suppes' demand will be satisfied if it is shown that $\mathscr{S}$, the axioms characterizing $\mathscr{S}$, and the part of the world thereby modeled can be studied and characterized without appeal to concepts of quantitative probability.

Thus, the various conditions of adequacy which I shall advance work *together* to ensure that a philosophical theory of probability which satisfies them all will be adequate. Indeed, just as satisfaction of formal adequacy will play an important role in ensuring the testability of a theory in connection with admissibility, so satisfaction of interpretation/idealization will play an important role in testing whether or not the theory has adequately identified the intended concept of probability, a concept in virtue of which conceptual adequacy may be satisfied. Also in the second section of this paper, we shall see the connection between the condition of interpretation/idealization and Salmon's criterion of ascertainability, according to which it must be possible, in principle, to ascertain the values of probabilities.

From third section to sixth section of this paper, I shall examine instances of what I take to be the four main kinds of interpretations of probability according to which probability is objective, or physical, with an eye towards the extent to which they satisfy the conditions of adequacy elaborated in the first two sections. I shall examine the actual limiting frequency conception, attributed to Von Mises (1957, 1964) and Reichenbach (1949), the hypothetical limiting frequency view as formulated by Kyburg (1974, 1978), the "long run" construal of Popper's (1957, 1959) propensity interpretation, and Fetzer's (1971, 1981) "single case" propensity view. All of these views can be formulated in such a way that they satisfy the conditions of formal adequacy and admissibility. What I shall argue is that none of them satisfies *both* the condition of conceptual adequacy *and* the condition of interpretation/idealization and that to the extent that they satisfy one of these two conditions, they fail to satisfy the other. I shall argue that as far as conceptual adequacy goes, the theories rank from better to worse roughly in the following order: single case propensity, long run propensity, hypothetical limiting frequency, actual limiting frequency. And I shall argue that with respect to interpretation/idealization, these theories rank roughly in the opposite order.

It is perhaps worth noting that this general kind of tension between the satisfaction of two desiderata of adequacy is, of course, not new in philosophy, nor even in the philosophy of probability. Fetzer (1974) has noted a tension between satisfaction of "epistemological criteria" (on which actual frequency conceptions of probability seem to be preferable to a single case propensity account) and "systematic considerations" (on which the propensity interpretation is preferable). In the philosophy of mathematics, Benacerraf (1973) argues that no adequate account of mathematical truth has allowed for an adequate account of mathematical knowledge, and vice versa – i.e., roughly, that reasonable theories of mathematical *truth* leave it unintelligible how we can obtain mathematical *knowledge*, while reasonable *epistemologies* fail to show how the suggested "truth conditions" are really conditions of *truth*.

While my condition of interpretation/idealization is more methodological than epistemological in character, these tensions are of the same general kind as the one I shall argue is present in the philosophy of objective probability. Perhaps closer to the tension I shall try to characterize in objective probability theory is one that can be found in the philosophical foundations of modal logic. While the analysis of possible worlds as maximally consistent sets of sentences makes the conception of possible worlds very *clear*, that conception is clearly also *theoretically inadequate* as a result, in part, of the limited expressive power of any available language. On the other hand, the analysis of possible worlds as, say, "ways the world could have been", while perhaps closer to the *theoretically intended* conception, would seem to be *methodologically unsound*, in that it would render the usual analysis of possibility and of counterfactuality in terms of possible worlds circular.

Actual Relative Frequencies

Although the finite relative frequency interpretation of probability, endorsed by Russell (1948) and mentioned favorably by Sklar (1970), has been forcefully criticized by many philosophers as being conceptually inadequate in several important respects, its basic features are relatively simple, and it will serve well as an example in terms of which the conditions of interpretation/idealization and formal adequacy (whose satisfaction is independent of satisfaction of the conceptual adequacy requirement) can be explained. On this interpretation, roughly, the probability of an attribute A in a reference class B is the relative frequency of occurrences of A within B, where A and B are the *finite* classes of *actual* occurrences of events of the relevant kinds.

To be more precise about the interpretation, we may define finite relative frequency structures (*FRF*-structures) $\mathscr{C}$ as follows. Where E is any finite class and F is the power set of E (i.e., the set of all subsets of E) and # is a function which assigns to any member of F its cardinality (i.e., the number of its elements), $\langle E, F, \#\rangle$ is an *FRF-structure*. (Alternatively, F may be *any* Boolean algebra of subsets of E.) Thus, an *FRF*-structure is any triple $\langle E, F, \#\rangle$ that satisfies certain axioms, which axioms will guarantee that F is 2^E and that # is the cardinality function. Such a structure is an example of a characteristic structure $\mathscr{C}$, alluded to in the first section, where *FRF*-structures are (part of) what finite relative frequentists might use to get their interpretation of probability to satisfy the condition of formal adequacy.

To complete the demonstration that the finite relative frequency theory satisfies the formal adequacy requirement, we show that certain features of an *FRF*-structure can be *represented* by a structure $\langle \mathscr{B}, P\rangle$, where $\mathscr{B}$ is a Boolean algebra and P is a probability function on $\mathscr{B}$. For an *FRF*-structure $\langle E, F, \#\rangle$, simply let $\mathscr{B}$ be F (alternatively, *any* Boolean algebra of subsets of E) and, for $A, B \in \mathscr{B}$, let $P(A) = \#(A)/\#(E)$ and $P(A/B) = \#(A \cap B)/\#(B)$. From the axiomatization of $\langle E, F, \#\rangle$, it can easily be shown that P is a probability on $\mathscr{B}$, i.e., that P satisfies probability axioms, where the arguments of P are the elements of $\mathscr{B}$. Thus, the characteristic kind of formal structure $\mathscr{C}$ for the finite relative frequency interpretation – *FRF*-structures – has been characterized, and it has been shown how certain features of an *FRF*-structure can be represented in terms of a probability on an appropriate structure determined (at least in part) by the *FRF*-structure. And it is just these two things – the *definition of the characteristic kind of* structure with the capacity to yield *probabilistic representation* – that must be given for the condition of formal adequacy to be satisfied.

Of course the specification of the characteristic kind of structure and the probabilistic representation of certain features of such structures does not by itself constitute an appropriate *interpretation* of probability. For these may be just abstract mathematical entities, where objective probability is supposed to apply to the physical world. In order to complete the interpretation, therefore, both the "intended model" – an *instance* of the characteristic *kind* of structure – and the intended interpretation of the intended model must be specified. For the finite relative frequency interpretation, this may be done as follows. An *FRF*-structure $\langle E, F, \# \rangle$, is the *intended model* if there is a one-to-one correspondence between E and some set of events (or trials) in the physical world.

Of course the intended model may be relativized to "local contexts", where there is a one-to-one correspondence between E and the set of events of "local interest", e.g., the set of all throws of dice, or of a particular die, or the set of all married American men who apply for life insurance at the age of 50. Then the sets in F correspond to the relevant properties, e.g., role of a die coming up "6", a person's dying before the age of 75, etc. The *intended interpretation* of the intended model is just the one-to-one correspondence between the elements of E and the relevant events in the world, and thus between the sets in F and corresponding properties, where # is interpreted as the cardinality function on the interpretation of F.

Thus, the condition of interpretation/idealization is satisfied when the intended characteristic structure (the intended model) and the theory's interpretation of the intended model are both given. The reason why the condition of interpretation/idealization is so-called is that it concerns specification of the intended model of the theory and the intended relation between that structure and the world, where (i) the relevant features of the world are an *interpretation* of that structure and (ii) the (perhaps abstract) entities, relations, functions, etc., of that structure, together with the structure's axiomatization, serve as an *idealization* of the relevant part of the world.

Talk of "*the* intended model", of course, is not meant to imply that there is, literally, *exactly one* such structure: rather, there is supposed to be (for the global context and for each local context) just one structure *modulo isomorphism*, where isomorphic structures are identified with one another. Also, I suppose it would be possible to collapse the two parts of interpretation/idealization – i.e., the specification of the intended model and the establishment of an interpretation of the model – into just one part, where, for the finite relative frequency interpretation, E is identified with the relevant set of events or trials in the world and the sets in F are identified with the relevant properties, construing properties extensionally as sets. But it is nevertheless worthwhile to distinguish conceptually between the role of the intended model (a structure whose constituents may be abstract mathematical entities and whose role is, in part, to show satisfaction of the condition of formal adequacy) and the role of the (ideally, physical and observable) entities in the world which the components of the intended model are interpreted as – the things to which probability is supposed to apply. For, in the case of the finite relative frequency interpretation, it seems that sets are indeed a rather crude *idealization* of properties which works well for the purposes of that interpretation. And in the second place, for other interpretations of probability (e.g., decision theoretic foundations of rational subjective probabilities, on which see, e.g., Eells 1982), both the *constituents* of the intended model and the *axiomatization* of the intended model are quite clearly very crude idealizations of the real-world entities and phenomena which they are supposed to model. The distinction in question is analogous to the distinction between two kinds of interpretations of Euclid's axioms for geometry: 'point' and 'line' can be interpreted *abstractly* as mathematical points (e.g., as pairs of real numbers) and abstract mathematical lines (e.g., sets of mathematical points $\langle x,\ y\rangle$ that satisfy $mx+b=y$ for some real numbers m and b); or they could be interpreted *physically* as physical points and physical lines (e.g., the possible paths of light rays). Similarly, the probability function may be interpreted abstractly as a function on an abstract intended model $\mathscr{S}$, and then also physically as a function on the features of the world modeled by $\mathscr{S}$, *via* the connection between those features and $\mathscr{S}$ established by satisfaction of the condition of interpretation/idealization.

Leaving aside for now the question of the conceptual adequacy of the finite relative frequency theory of probability (which will be discussed in the next section along with the conceptual adequacy of the actual limiting frequency interpretation), note how satisfaction of the conditions of formal adequacy and interpretation/idealization work together to ensure satisfaction of admissibility and to effect what Suppes (1974) calls "systematic definiteness" of the interpretation. Let A and B be any properties – or sets of events – in whose probabilities we may be interested, or in probabilistic relationships between which we may be interested. On the finite relative frequency theory, we must, guided by a local or global context, construct the intended model $\langle E,\ F,\ \#\rangle$, where there are sets, say A' and B', corresponding to A and B, such that $A',\ B' \in F$. According to the rules given in the probabilistic representation part of the satisfaction of the condition of formal adequacy, we get the structure $\langle \mathscr{B},\ P\rangle,\ P$ being a probability on $\mathscr{B}$, where $\mathscr{B}$ includes both A' and B'. Then, on the finite relative frequency interpretation of probability, the probability of A, and of A given B – in symbols, $prob(A)$ and $prob(A/B)$ – are just $P(A')$ and $P(A'/B')$. Also, we know that the interpretation satisfies the condition of admissibility, since P satisfies probability axioms; this is ensured by the interpretation's satisfaction of formal adequacy.

It should be clear that the conditions of formal adequacy and of interpretation/idealization must be satisfied by any satisfactory interpretation of probability. As to formal adequacy, an interpretation of probability is, after all, supposed to be an interpretation of the function symbol that appears in some axiomatization of probability, and it is difficult to see how it could possibly be shown that a purported interpretation of that symbol satisfies the axioms unless some kind of formal structure $\mathscr{B}$, e.g., a Boolean algebra, is provided by the theory. And if probability is to be some feature of the world – some kind of physical probability or even degree of belief – then the structure $\mathscr{B}$ cannot come from just anywhere: it must be related to the world in some appropriate manner. Some features of the world that are understandable, or at least capable of being studied, independently of probability must be identified. And for these features to be systematically related to the structure $\mathscr{B}$, these features must first be idealized and abstractly represented in terms of some structure $\mathscr{S}$ characteristic of the interpretation of probability in question, so that one can demonstrably infer that $\mathscr{B}$, together with a probability P on $\mathscr{B}$, represents the appropriate features of the intended model $\mathscr{S}$, and thus, indirectly, the appropriate features of the world. The general picture is as indicated in Fig. 1, where the concepts, objects and phenomena appropriate to some familiar interpretations of probability other than the finite relative frequency interpretation are indicated. Note that the brackets on the left overlap, indicating that the specification of the *characteristic kind* of structure pertains to formal adequacy, where identification of the *intended* model pertains to interpretation/idealization.

It is of some interest to compare the interpretation/idealization requirement with Salmon's "*criterion of ascertainability*":

> This criterion requires that there be some method by which, in principle at least, we can ascertain values of probabilities. It merely expresses the fact that a concept of probability will be useless if it is impossible in principle to find out what the probabilities are. (1967, p. 64)

The condition of interpretation/idealization is intended, in part, to capture the idea that probabilities should be ascertainable, but in a weaker sense than Salmon's criterion seems to require. The condition is only intended to ensure that probability statements have "empirical interpretation" – or "empirical content" – in a weaker sense similar to the one assumed by some versions of the hypothetico-deductive model of science. Consider Fig. 1. The entities of the top box can be thought of as *observable* (or, at least "*pre-theoretical*", i.e., "*pre-probability-theoretical*") *entities*, and the laws that govern them as *lawlike empirical generalizations* expressed in terms of an *observational* (or "*pre-probability-theoretical*") *vocabulary*. The concept of probability, as it figures in the bottom box, can be thought of as a *theoretical concept* of the philosophical theory of probability in question, while the probability axioms, together with the mathematical principles that relate the probability concept to the intended structure $\mathscr{S}$, can be thought of as the *theoretical* or *internal principles* of the philosophical theory of probability, which principles are expressed in terms of a *theoretical vocabulary*. And finally, the principles of interpretation/idealization, symbolized by the arrows between the top and middle box, can be thought of as *bridge principles* which relate the observable (pre-theoretical) entities of the top box to the mathematical entities of the middle box, in terms of which the theoretical concept of probability of the bottom box is characterized *via* what I have been calling "probabilistic representation".

On the hypothetico-deductive model of science, of course, the bridge principles are supposed to function as *contextual* or *implicit* or *partial definitions* of theoretical terms: they need not *completely* specify the meanings of the theoretical terms.

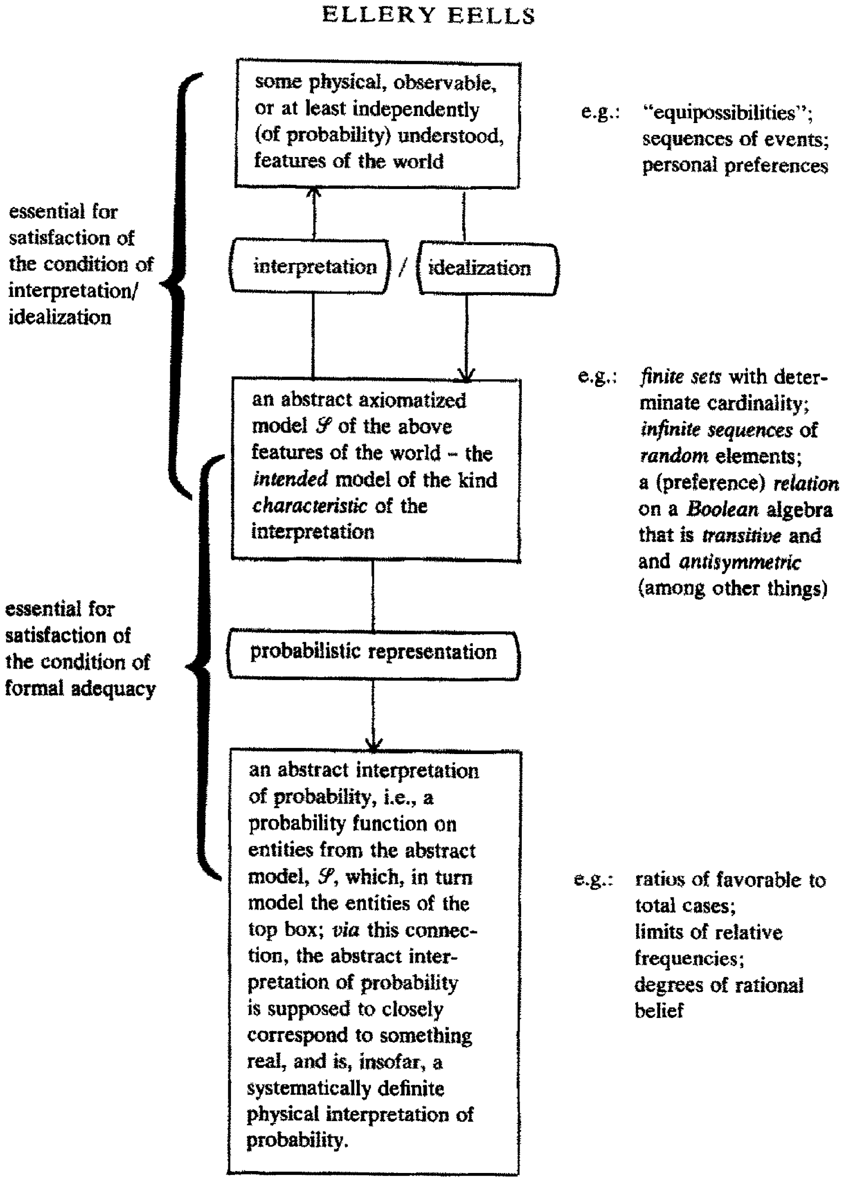


Fig. 1 Probability Models as Scientific Theories

Given this, and given that the principles of interpretation/idealization of a philosophical theory of probability are supposed to function much as the bridge principles of the hypothetico-deductive model of science, it should not be surprising if some interpretations of probability will not be able to specify empirical methods for ascertaining, even in principle, exact numerical values of all probabilities. The model of philosophical theories of probability presented in this section does not require that philosophical theories of probability be ascertainable in the stronger sense which Salmon seems to require. This would seem to be a virtue of initial neutrality between various philosophical theories of probability, for if the stricter version of ascertainability were insisted upon, it would seem that certain theories – certain propensity and dispositional accounts as well as certain Bayesian theories[1] – would be ruled out from the outset.

In concluding this section, I would like to emphasize again some important connections between the four conditions of adequacy suggested above and how they work together to ensure that an interpretation of probability that satisfies them all will be an adequate theory. We have seen that satisfaction of the condition of formal adequacy is required to demonstrate satisfaction of admissibility. Both formal adequacy and interpretation/idealization are required to show that the phenomena *in the world* upon which a given interpretation of probability focuses are indeed *probabilistic* phenomena in the sense that abstract probability theory applies to them. And finally, satisfaction of interpretation/idealization is supposed to identify precisely the intended concept and supply empirical (or, at least, "pre-probability-theoretical") content – *cognitive significance* for positivists – for the *conception* of probability offered, without which it would seem that satisfaction of the condition of conceptual adequacy would be empty.

Finite Relative Frequencies

The actual limiting frequency view of probability is a generalization of the finite relative frequency theory which is supposed to be applicable even if the relevant classes have infinite cardinality. My presentation of the characteristic kind of structure for the actual limiting frequency view will roughly follow that of Suppes' (1974), except for notation. An *ALF-structure* is any triple $\langle E,\ F,\ \#\rangle$ such that E is a sequence (by which I shall understand any function whose domain is the natural numbers,[2] where I shall write 'E_i' rather than '$E(i)$', for the value of the function with argument i), F is the power set of the range of E (i.e., the set of all subsets of the set of possible values of E), and # is the binary function such that for any natural number n and element A of F, $\#\,(A, n)$ is the cardinality of the set $\{E_i : i \leqslant n$ and $E_i \in A\}$. (Alternatively, F may be taken to be *any* Boolean algebra of subsets of the range of E.) Probabilistic representation proceeds as follows in terms of relative frequencies. The *relative frequency of a set* $A(\in F)$ *in the first n terms of E* is defined to be $\#(A, n)/n$. The *limiting relative frequency* of A in E is defined to be the limit of the relative frequency of A in the first n terms of E as n approaches infinity (if the limit exists), where the limit of a real-valued function $f(n)$ as n approaches infinity, $\lim_{n\to\infty} f(n)$, is defined to be that number r (if any) such that for every $\in > 0$, there is a natural number $N_\in$ such that for all $n > N_\in$, $|f(n) - r| < \in$. Now let $\mathscr{B}$ be any Boolean algebra of subsets of the range of E such that the limiting relative frequency of every element of $\mathscr{B}$ in E exists. Then the structure $\langle \mathscr{B},\ P\rangle$ is a probabilistic representation of the *ALF*-structure $\langle E, F, \#\rangle$, where for any A and B in $\mathscr{B}$,

$$P(A) = \lim_{n\to\infty} \frac{\#(A, n)}{n},$$

and

$$P(A/B) = \lim_{n\to\infty} \frac{\#(A\cap B, n)}{\#(B, n)} = \frac{P(A\cap B)}{P(B)}.$$

An alternative approach (the one I shall have in mind in the sequel) would include in the axiomatization of *ALF*-structures the stipulation that all limiting frequencies of elements of F exist (Axiom of Convergence, or Limit Axiom), so that the set F could not in general simply be the power set of the range of E. This has the effect that for any *ALF*-structure $\langle E, F, \#\rangle$, there would be a uniquely characterizable probabilistic representation: the function P (defined above) on F itself. For conceptual adequacy, one might also want to include an Axiom of Randomness (Principle of the Excluded Gambling System), such as Von Mises', in the axiomatization of *ALF*-structures. In any case, it should be clear that the actual limiting frequency view of probability is admissible and formally adequate.[3]

For satisfaction of the condition of interpretation/idealization, the actual limiting frequency interpretation must specify an intended model, a particular *ALF*-structure, for the "global" context or any given "local" context. For the global context, one may insist on one-to-one correspondences between the natural numbers n and *times*, between the range of E and the set of past, present and future (actual) events, and between F and the relevant attributes. Then perhaps the natural number 1 could correspond to the "first" time, an element E_n of the sequence would correspond to what happens at the nth time, and so on. This assumes, of course, that the universe is temporally finite in the pastward direction, which assumption could be gotten around, though, by not insisting that E orders events temporally. That time is not dense is also assumed here. More plausibly, however, the limiting frequency view is applicable to local contexts, where the range of E is a set of events of local interest, e.g., tosses of a particular die, or of all dice, or human births, or applications for life insurance.

For the limiting frequency view of probability, unlike the finite relative frequency view, probability attaches to *ordered* sets of events, i.e., sequences of events whose order is given by the underlying sequence E of *all* events relevant to the local or to the global context. And it is clear that a solution to the problem of interpretation/idealization must specify an intended order in which the relevant events are to be taken. This is because, as long as there are infinitely many elements of E (or of a set B) that are elements of a set A and also infinitely many that are not elements of A, then the limit of the relative frequency of A's in E (or in B) could be any number whatsoever between 0 and 1, inclusive, depending on the order in which the events are taken: in this case, *the order completely determines the probability*. Supposing that this *is* a problem for the actual limiting frequency view, then would it be a problem in connection with satisfaction of conceptual adequacy or a problem in connection with satisfaction of interpretation/idealization? The answer depends on the nature of the (perhaps more or less vague) *conception* of probability, whose theoretical adequacy must be certified if the theory is to satisfy the condition of conceptual adequacy, and which must be made precise and given empirical content if the theory is to satisfy the condition of interpretation/idealization. It is the job of interpretation/idealization to make that conception clear and precise (which, for the purposes of this paper, I am assuming may be theoretically adequate, or not, independently of its clarity and precision). So if the conception were vague and noncommittal with respect to the order of the events, the problem described above would be one for interpretation/idealization: a solution to the problem of interpretation/idealization must then justify one order among the many possible as the intended one. But if the conception of probability offered were clear with respect to the intended order of events, then the actual limiting frequency theory would be conceptually adequate, or not, in part to the extent to which the order to which the conception is committed is itself theoretically justified. Since in all of the discussions of the application of probability construed as limiting frequency (which I have seen, at any rate) it is clear that the intended order of events is their temporal order, I shall assume that the *conception* of probability is clear about the intended order of the relevant events: it is their temporal order. Thus, the problem under discussion is not a problem for interpretation/idealization. Indeed, it seems that the actual limiting frequency interpretation fares quite well on the condition of interpretation/idealization: all of the relevant concepts – that of events, of temporal order, of cardinality, of sequences, of sets, of limits, etc. – are either fairly well understood or at least such as can be studied and understood independently of the concept of probability.

So the actual limiting frequency view (as understood for the purposes of this paper) will be conceptually adequate, or not, in part to the extent to which using the temporal order of events in calculating limiting frequencies is theoretically justified. I can think of no reasons in favor of or against using the temporal order rather than any other order, but the absence of reasons in either direction might itself suggest the argument that using the temporal order is *arbitrary*. Also, note that one effect of always using *one* order, such as the temporal order, of events is to make the probability of an attribute A – or the probability of an attribute A within a reference class B – invariant over time, where (although I have no natural suggestions along these lines) perhaps one way of accommodating the intuitive possibility that probabilities might *change* over time would be to use different orders of the relevant events at different times. The fact that, intuitively, probabilities seem to change over time (e.g., the probability of a person living more than 60 years has, intuitively, changed over time) – while it seems that, on limiting frequency conceptions, all probabilities are fixed for all time – will be recognized as a version of the reference class problem, or of the problem of the single case.

The frequentist response, of course, is that $P(A/B)$ – where B is the class of incidents of human births and A is the class of incidents of a being living more than 60 years – has *not* changed, but that different reference classes are appropriate for assessing probabilities with predictive significance at different times. Let C be the class of incidents of human births where the person will enjoy the results of modern medical advances throughout his life. Then $P(A/B \cap C)$ is one probability and

$P(A/B \cap \bar{C})$ is another, where the first is more appropriate for predictive purposes today and the second would have been more appropriate, say, during the Dark Ages.

The problem of the single case for the actual limiting frequency theory of probability is usually formulated as follows. On the frequency view, probability attaches to (ordered) *collections* of events; but we are sometimes interested in the probability that, e.g., the *next* event will exemplify some attribute. The problem of the single case, then, asks how the frequency interpretation can apply to single events. And the general form of proposed solutions to the problem is to give a rule for choosing an appropriate reference class to which the single event belongs and to say that the probability of the relevant attribute within that reference class should be transferred to the single event in question. Hence, the problem is sometimes also called 'the problem of the reference class'.

I do not believe that there is an adequate solution to the problem of the reference class for the actual limiting frequency view of probability. And in light of the fact that we would ideally like an explication of physical probability to have predictive and explanatory significance for single events (which may occur irreducibly probabilistically) and to have significance in connection with decision making in individual decision situations, this constitutes a serious limitation to the *conceptual adequacy* of the theory. Without going into much detail, let me summarize some of the considerations that have been, or could be, brought to bear against two proposed solutions to the problem.

Reichenbach's (1949, p. 374) solution to the reference class problem, of course, was to choose "*the narrowest class for which reliable statistics can be compiled*". Thus, in relation to the above example, to assess the probability that a particular individual will live more than 60 years, it is better to use one of $B \cap C$ and $B \cap \bar{C}$ as the reference class than just B, depending on which class the individual in question belongs to, provided that reliable statistics – with respect to life span – can be compiled for the two classes. It is also part of Reichenbach's (1949, p. 374) solution that we do not narrow a class with respect to another class when the second class is "known to be irrelevant"; that is, if class D is known to be such that $P(A/B \cap C \cap D) = P(A/B \cap C)$, then we should not favor the use of the narrower class $B \cap C \cap D$ over use of $B \cap C$. Note two features of this solution. First, there is a subjectivist element in the solution, in that the choice of reference class depends on the reliability of our knowledge of the relevant statistics. Second, as Salmon (1971, p. 41) has pointed out, it is often the case that the more reliable the statistics, the broader the reference class must become, and the narrower the reference class, the less reliable the statistics become. This is in part because, for classes A, B, and C, $P(A/B \cap C)$ is not a function of – cannot be calculated from only – $P(A/B)$ and $P(A/C)$.

Now Reichenbach (1949, p. 375) insisted, of course, that, literally speaking, probability applies only to sequences. And in connection with the single case, he says,

> We are dealing here with a method of technical statistics; the decision for a certain reference class will depend on balancing the importance of the prediction against the reliability available.

In a similar vein, Salmon suggests that, plausibly, Reichenbach

> was making a distinction similar to that made by Carnap between the principles belonging to inductive logic and methodological rules for the application of inductive logic. The requirement of total evidence, it will be recalled, is a methodological rule for the application of inductive logic. Reichenbach could be interpreted as suggesting analogously that probability theory itself is concerned only with limit statements about relative frequencies in infinite sequences of events, whereas the principle for selection of a reference class stands as a methodological rule for the practical application of probability statements. (1971, p. 41)

But one may nevertheless insist that an account of physical probability is conceptually inadequate unless, on the account, probability applies objectively to single events, which is not implausible if one thinks, in the first place, that ideally what one would like to know in particular decision problems is the controlling objective probabilities, and, in the second place, that there can be correct and complete statistical explanations of particular events that may occur irreducibly probabilistically. Such an account of probability cannot let the values of probabilities depend on the incomplete state of our knowledge, which a purely methodological account of single case probabilities must require.

But, in light of the distinction Salmon suggests Reichenbach might have had in mind, perhaps the actual limiting frequency view can be elaborated in such a way as to apply objectively to the single case. To get around being forced into a *methodological* context by the *incompleteness* of our state of knowledge, we simply envisage, for the purpose of remaining in a nonmethodological context of *explicating* single case objective probabilities, a hypothetical state of *complete* knowledge with respect to the relevant facts. This would seem to have the promise of eliminating the subjectivity of Reichenbach's solution, eliminating the practical conflict between reliability of statistics and narrowness of the reference class, while preserving the frequency conception of probability. I think that part of Fetzer's (1977; see also his 1981, pp. 78–86) paper "Reichenbach, Reference Classes, and Single Case 'Probabilities'" can be viewed as following up this idea and showing that it can only result in a trivialization of single case probabilities on the frequency view, where all such probabilities will turn out to be either 0 or 1.

Following up Reichenbach's (1949, pp. 375–376) general ideal that "the probability will approach a limit when the single case is enclosed in narrower and narrower classes, to the effect that, from a certain point on, further narrowing will no longer result in noticeable improvement", Fetzer defines an *ontically homogeneous reference class with respect to an attribute A and a trial* (*single event*) x (roughly) as a class B such that $x \in B$ and for all $B' \subseteq B,\ P(A/B) = P(A/B')$, and he suggests that on Reichenbachian principles, the appropriate reference class for x relative to A would be some ontically homogeneous reference class with respect to A and x. But which one? Presumably, the appropriate one would be the first one that one "reaches" in successively narrowing some initial candidate with respect to which Fetzer calls *permissible predicates* (i.e., predicates which are permissible for use in the description of a reference class): predicates which *do not imply the presence or absence of the relevant attribute*, which *are not satisfied by at most a finite* *number of things on logical grounds alone*, and which *are satisfied by at least one thing* (presumably the single event in question). And it is then argued, on the basis of the principle

If x and y are different events, then there is a permissible predicate F such that Fx and $\bar{F}y$,

that the set of permissible predicates satisfied by a single case x will not all be satisfied by any other single case y, so that the "appropriate" reference class turns out to be just $\{x\}$ and all single case probabilities turn out to be either 0 or 1, depending on whether the single event in question lacks or has the relevant attribute.

Note that it seems that Fetzer's argument assumes that conjunctions of permissible predicates are themselves permissible. This, of course, would need some argument, for it is obviously possible for two predicates, neither of which is satisfied by at most a finite number of events on logical grounds alone, to have a conjunction which is. For example, let F be the predicate "is the sinking of the *Titanic* or happens before the year 1900" and let G be the predicate "is the sinking of the *Titanic* or happens after the year 1900". As another example, let F specify just spatial coordinates and G a time. Perhaps the intent, however, is that permissible predicates must all be *dispositional* predicates of some kind, where his theory of dispositions (see especially his 1981, pp. 160–161 and 190–192) would somehow ensure that conjunctions of permissible predicates will be permissible. In any case, Fetzer's *main* point still holds, namely that on actual limiting frequency conceptions of probability, it is impossible to distinguish between factors that are statistically relevant *because of a "real causal" connection* and those which are statistically relevant *purely by coincidence*. And whether or not it is always possible to describe a given event *uniquely* in terms of permissible predicates is not so much at issue.

Suppose that in fact some event x is the only event in the course of the world's actual history that satisfies each of the predicates $F_1, \ldots, F_n$, each assumed to be permissible in some correct sense of 'permissible'. Then the actual limiting frequency of the relevant attribute, say A, in the class, say B, of individuals that satisfy each of $F_1, \ldots, F_n$ is either 0 or 1, depending on whether x lacks or has attribute A. And it is surely possible that, for any i between 0 and n, there are *many* events which satisfy each of $F_1, \ldots, F_{i-1}, F_{i+1}, \ldots, F_n$. And, where, for each such i, B_i is the reference class of events that satisfy each of $F_1, \ldots, F_{i-1}, F_{i+1}, \ldots, F_n$, it is clearly also possible that the $P(A/B_i)$'s (on the actual limiting frequency interpretation) all differ from each other and from $P(A/B)$ (on the actual limiting frequency interpretation). And all this is consistent with *none of the F_i's having anything physically to do with the presence or absence of A in x*: it just happens that x is the only event in the course of the world's actual history that satisfies each of the F_i's. And the fact that it is possible that there be *just one* event is not so much to the point. Suppose that in fact, in the entire course of the world's actual history, there will be only three instances of a well-balanced coin being fairly tossed when the moon, Mars and Halley's comet are in close opposition. Then the actual limiting frequency of a given such event resulting in tails-up will be either 0, $\frac{1}{3}$, $\frac{2}{3}$, or 1, yet, intuitively, it is absurd to conclude that the celestial configuration described introduces a *physical bias* into the trial, just because removal of any of the three factors (moon in opposition, Mars in opposition, Halley's comet in opposition) yields a limiting relative frequency of about $\frac{1}{2}$. Intuitively, it would seem that the *probability* of tails on any of the three tosses is (about) $\frac{1}{2}$, though of course the *actual frequency* of tails in the circumstances described is artificially limited to being one of the four values specified above.

Relative frequentists could respond to such examples in a number of ways. They could, for example, say that in order for an actual relative frequency to be the true single case probability, one must use a large enough reference class for which reliable statistics are available. Thus, Salmon (1971) urges that, instead of using the *narrowest* such reference class, we should use the *broadest* homogeneous such reference class to which the single case in question belongs, where a *homogeneous reference class for an attribute A* is defined to be a class for which it is impossible to effect a statistically relevant partition (with respect to A) without already knowing which elements of the class have attribute A and which do not. This takes seriously Reichenbach's (1949, p. 374) idea that, "Classes that are known to be irrelevant for the statistical result may be disregarded". But still, it would seem that as long as we are dealing with actual relative frequencies, such a partition could be statistically relevant just as a matter of coincidence, as when we partition the class of tosses of honest coins by whether or not they occur when the moon, Mars and Halley's comet are in close opposition. See Fetzer (1977, pp. 199–201; and 1981, pp. 91–92) for another kind of criticism of Salmon's approach and for further discussion.

On the other hand, one may simply insist upon the use of some infinite sequence of events which are similar to the single event in question in all relevant (causal) respects. But, first, this would require an explication of causal relevance prior to an explication of probability, where this would render circular recent attempts to explicate causality in terms of probability relations.[4] And second, even if this could be done, it is possible that there might be, say, only two or three events in the course of the world's actual history that are similar to the single event in question in all relevant respects.

A frequentist may respond to this last difficulty along the following lines, as Salmon (1979, pp. 11–12) has suggested that Reichenbach would (see his 1949, §34). Instead of considering the three tosses discussed above as members of some *actual* (and thus possibly finite) sequence, consider them as members of the sequence of tosses that *would* exist *were* we to toss the coin infinitely many times under the same circumstances, and then ask what the limiting relative frequency of tails would be in this sequence. Of course this is to abandon the idea that probability should be explicated in terms of sequences of *actual* events. In the next section, we look at the *hypothetical* limiting frequency interpretation, which attempts to specify appropriate principles for extending actual finite sequences (e.g., single element sequences) to hypothetical infinite sequences.

But as to the *actual* limiting frequency interpretation of probability, it seems correct to conclude that, while the theory is basically *adequate as far as interpretation/idealization* (and formal adequacy and thus also admissibility) *goes*, it is *conceptually inadequate* in that, although it may be argued to have appropriate predictive significance, it is incapable of characterizing the difference between genuine physical connections and merely historical coincidences and is, largely for this reason, incapable of applying appropriately to single events.

Hypothetical Relative Frequencies

In order to accommodate some of the difficulties discussed above in connection with actual relative frequency conceptions of probability, a hypothetical limiting frequency conception may be advanced, according to which the probability of an attribute A – or the probability of B's being A's – is equal to the limiting frequency of A in a hypothetical infinite extension of the actual (finite) sequence of events – or a hypothetical infinite extension of the actual (finite) sequence of B's. Thus, $P(A)$ – or $P(A/B)$ – is supposed to be what the limiting frequency of A *would be* – or what the limiting frequency of A in the sequence of B's *would be* – if the world's history *were* infinite – or if the sequence of B's *were* infinite. This is the basic *conception* of probability on hypothetical limiting frequency views, which conception must be given precision in a solution to the problems of formal adequacy and of interpretation/idealization if the interpretation is to be adequate.

Kyburg (1974, 1978) formulates semantics for the hypothetical limiting frequency view (which may be viewed as a solution to the problem of formal adequacy and part of a solution to the problem of interpretation/idealization) as follows. He begins with a first order language with identity and enough mathematical machinery to axiomatize the three place predicate S in such a way that '$S(A, B, r)$' is true in a model $M = \langle U, R\rangle$ (where R is a set of relations on U and functions on U, U^2, etc., and U contains at least the empty set, Ø) if and only if[5]

(i) B is an infinite sequence of sets, $B_1,\ B_2, \ldots$, where $B_i \subseteq B_j$ if $i < j$,
(ii) A is a set.
(iii) r is a real number.
(iv) $r = \lim\limits_{i\to\infty} \frac{\#(A\cap B_i)}{\#(B_i)}$.

The relations and functions of these models are assumed to be "compatible" in the sense that if $\langle U, R\rangle$ and $\langle U', R'\rangle$ are any two such models with $U \subseteq U'$, then (i) for every predicate symbol A of the language, $R(A) \subseteq R'(A)$ (where $R(A)$ and $R'(A)$ are the relations which R and R' assign to A) and (ii) for any k-place function symbol f of the language, if $x \in U^k$, then either $R(f)(x) = R'(f)(x)$ or $R(f)(x) = \emptyset$. (The reason for insisting that Ø is an element of every model is so that what would otherwise be a partial function may take Ø as a value where it would otherwise be undefined.)

The "actual world" is taken to be a particular model, $M^* = \langle U^*, R^*\rangle$. A *future model* is any model in which every (actually) true observation sentence pertaining to times up to the present is true. And a *lawful model* is any model in which all (actually) true universal (nonstatistical) physical laws are true. A model $M' = \langle U', R'\rangle$ is an *extension* of a model $M = \langle U, R\rangle$ if $U \subseteq U'$ and the relations and functions in R are the restrictions to U of the relations and functions in R'. Finally, for a term B interpretable as a sequence, a model M' is a *B-maximal extension* of a model M if M' is an extension of M and either B is infinite in M' or no extension of M' extends B. Thus, while all lawful future worlds may have finite histories, B-maximal extensions of such worlds may have infinite histories. Finally, truth conditions for hypothetical limiting frequency statements, '$P(A/B) = r$', are given as follows:

> '$P(A/B) = r$' is true (in the actual world) just in case '$S(A, B, r)$' is true in every (or "almost every") B-maximal extension of every lawful future world.

This is not *exactly* the same semantics for hypothetical relative frequency statements as that given by Kyburg – indeed, he considers several variations – but it is close and captures all the features of the hypothetical limiting frequency interpretation which I wish to discuss.

I shall structure the discussion of the hypothetical limiting frequency view around the conditions of adequacy discussed in the first two sections of this paper. As to conceptual adequacy, the rough conception of probability offered was presented above, and all I have to say about the conceptual adequacy of the view is that it seems clearly superior to the actual limiting frequency conception in that it deals with the possibilities that the actual history of the world is *finite* (where the actual frequencies may, in some cases, be *ratios of small numbers* that don't faithfully represent the relevant features of the physical world) and that it may exhibit *coincidences*. It deals with the first possibility by envisioning hypothetical *infinite extensions* of the actual world's history and with the latter possibility by considering *many* extensions of the actual world's history, where, presumably by invoking the law of large numbers idea, this is taken to accommodate actual world coincidences of a global character (but see below on this). The solutions to the problems of formal adequacy and interpretation/idealization are then supposed to clarify this rough conception.

Recall that a solution to the problem of formal adequacy is supposed to identify the *characteristic kind of structure*, some features of which can be given *probabilistic representation*. I suggest that a characteristic structure for the hypothetical limiting frequency interpretation be understood to be of the following form (variants are possible, as discussed below): an *HLF-structure* is an (appropriately axiomatized – see below) sextuple, $\langle \mathscr{L}, \mathscr{M}, M^*, F, L, E \rangle$, where $\mathscr{L}$ is a first-order language with axioms (at least for the three-place predicate S), $\mathscr{M}$ is a set of models $M = \langle U, R \rangle$ of the kind defined above, $M^* \in \mathscr{M}$, F and L are subsets of $\mathscr{M}$, and E is a relation in $\mathscr{L} \times \mathscr{M} \times \mathscr{M}$. Then truth conditions for hypothetical limiting frequency statements could alternatively be given as follows:

> '$P(A/B) = r$' is true (relative to a given *HLF*-structure) if and only if for every $M \in \mathscr{M}$ such that $F(M)$ and $L(M)$, '$S(A, B, r)$' is true in every (or almost every) $M' \in \mathscr{M}$ such that $E(B, M, M')$.

Also, if "every (or almost every)" could be made precise in an appropriate way, it would be possible to infer, in light of these semantics, a finitely additive probability space from a structure characteristic of the theory, thus explicitly satisfying the probabilistic representation part of the condition of formal adequacy as formulated above. Of course the intended model of the hypothetical limiting frequency interpretation would have M^* correspond in some appropriate way to the actual world, where F is the set of future worlds (as described above), L is the set of lawful worlds (as described above), and E is the relation such that $E(B, M, M')$ is true in the intended model if and only if M' is a B-maximal extension of M (as described above). And characteristic structures would be axiomatized in such a way as to guarantee that the purely formal relations between, e.g., worlds M and B-maximal extensions of M would hold, e.g., the axioms should imply that if $E(B, M, M')$ is true in a characteristic structure, then $U \subseteq U'$, where $M = \langle U, R \rangle$ and $M' = \langle U', R' \rangle$.

Before further investigating the formal adequacy of the hypothetical limiting frequency theory, it is worth pointing out that *HLF*-structures could have been construed differently. For example, L could be construed as a binary relation on $\mathscr{M}$, where $L(M, M')$ holds in the intended *HLF*-structure if and only if all the universal laws that hold in M also hold in M', and similarly for F. Or L could be thought of as a function from $\mathscr{M}$ to subsets of $\mathscr{L}$, where, in the intended model, $L(M)$ is the set of universal laws true in M. Then the set of lawful-relative-to-M models could be identified in the obvious way – and similarly for F.

As far as formal adequacy goes, it seems that the only unclarity in the hypothetical limiting frequency theory is in connection with the phrase "every (or almost every)". Which is it? Without at least a specification of which it is, the truth conditions for '$P(A/B) = r$' aren't definite, and it would not be possible to construct a finitely additive probability space from an *HLF*-structure in the light of the given semantics. Consider first the possibility of reading the phrase as "every". This would surely give us formal adequacy of the theory, but it would render *every* probability statement false *in the intended model* (i.e., given the intended meanings of 'future model', 'lawful model', etc.), thus rendering the interpretation inadequate in relation to interpretation/idealization. Consider, for example, statements of the form 'The probability of tails on a "fair" toss of this coin is r', and let us assume that this statement form – in symbols, '$P(A/B) = r$' – yields a true statement (intuitively) just when '$\frac{1}{2}$' is substituted for 'r'. Given the truth conditions for statements of hypothetical relative frequency stated above – and reading "every (or almost every)" as "every" – a statement '$P(A/B = r$' is true if and only if '$S(A, B, r)$' is true in *every* B-maximal extension of every lawful future world. But surely there is *some* such world in which '$S(A, B, 1)$' is true, i.e., in which the limiting relative frequency of tails is 1. And, as Skyrms says,

> On the hypothesis that the coin has a propensity of one-half to come up heads on a trial and that the trials are independent, *each infinite sequence of outcomes is equally possible. If we look at all physically possible worlds, we will find them all*, including the outcome sequence composed of all heads. (1980, p. 32)

That is, there is also *some* B-maximal lawful future world in which '$S(A, B, 0)$' is true. Thus, for *no* value of r is '$S(A, B, r)$' true in *all* B-maximal extensions of lawful future worlds. So, let us abandon the "every" reading of "every (or almost every)" and consider now the "almost every" reading.

How are we to understand "almost every" in a precise way? We surely cannot take it to mean "all but a finite number", for, in the coin tossing example of the previous paragraph, if there is one world in which '$S(A, B, r)$' is true for some value of r, then surely there are infinitely many such worlds. Fetzer and Nute (1979, 1980; see also Fetzer 1981, 56ff) have suggested the following way of making Kyburg's truth conditions precise on the "almost every" reading. Where $M_1, M_2, \ldots$, is an infinite sequence of B-maximal extensions of lawful future worlds,

$$\begin{aligned}&\text{'}P(A/B) = r\text{' is true (in the actual world } M^*\text{) if and only if}\\ &\lim_{k\to\infty} \frac{\#\{M_i : i \leqslant k \text{ and '}S(A,\ B,\ r)\text{' is true in } M_i\}}{k}\\ &= 1.\end{aligned}$$

Actually, this is a slight variant of Fetzer and Nute's suggestion, which is closer to Kyburg's formulation: the former assume that the worlds M_i are *themselves* future and lawful, despite their having *infinite* sequences of B's (more on this general idea below).

There are several difficulties with this proposal. First, there are, presumably, at least continuum many B-maximal extensions of any lawful future world, where, again, 'B' means 'this coin tossed' and 'A' means 'tails': for each infinite sequence of heads and tails, there is at least one B-maximal extension of any lawful future world, and infinite sequences of heads and tails can be identified with functions from the natural numbers into {heads, tails}, of which there are continuum many. So the natural question at this point is, "On what principles do we select the *denumerably* long sequence $M_1,\ M_2, \ldots$, of B-maximal extensions of lawful future worlds from the *nondenumerably* many such worlds?" Of course for any value of r, with $0 \leqslant r \leqslant 1$, there are infinitely many sequences of B-maximal extensions of any lawful future world such that, using any one of them, the truth conditions suggested will yield the truth of '$P(A/B) = r$'. This is because there are infinitely many infinite sequences of heads and tails for which the limiting relative frequency of tails is r, for any value of $r,\ 0 \leqslant r \leqslant 1$.

Now *part* of the problem of selecting an appropriate sequence of B-maximal extensions of lawful future worlds would be solved if plausible principles governing which B-maximal extensions of lawful future worlds should be *elements* of such a sequence could be provided. Perhaps we should require that such B-maximal extensions of lawful future worlds themselves be lawful. (It seems that part of the intent of the definition of 'extension' is that all extensions of future worlds will themselves be *future*, for it is natural to assume that the sequences in R of a world $\langle U, R\rangle$ are sequences of events taken in their temporal order, so that all B-maximal extensions of lawful future worlds will automatically be future worlds.) But it would require an argument to establish that *any* B-maximal extension of *any* lawful future world is lawful, if $R(B)$ is supposed to be *infinite* in any B-maximal extension $\langle U, R\rangle$ of any lawful future world: conceivably, the universal laws true of the actual world might imply that the world will have a finite history, as some cosmological models predict. Of course Kyburg's definition of B-maximal extension has a clause in it to handle the case in which there is no extension which extends a finite sequence B. But it is not explained *why* there may be no such extension in some cases, e.g., whether or not it could be a matter of physical law. And note that whether or not the actual world's history is finite as a matter of physical law should not, intuitively, control whether or not some probabilities have irrational values.

But even if plausible principles for selecting a denumerable *set* of B-maximal extensions of lawful future worlds could be given, there would remain the problem of *ordering* the models in this set so as to obtain the infinite *sequence* of models required for the truth conditions to be applicable. As long as there are infinitely many *not*-$S(A, B, r)$-worlds as well as infinitely many $S(A, B, r)$-worlds in a given set of B-maximal extensions of lawful future worlds, the truth of '$P(A/B) = r$' on the suggested truth conditions will depend on the particular *order* in which the set of B-maximal extensions is taken, for any value of r, for this order will determine whether or not the main sequence in the truth conditions converges to 1. Note also that if there are infinitely $S(A, B, r)$-worlds and infinitely many $S(A, B, s)$-worlds in a given set of B-maximal extensions of lawful future worlds, there will be infinitely many orderings of the set which will yield the truth of '$P(A/B) = r$' and also infinitely many that will yield the truth of '$P(A/B) = s$', and this conditional statement holds for *any* values of r and s.

Now none of these difficulties with Fetzer and Nute's suggestion is a deep one for the problem of formal adequacy. Instead of taking $\mathscr{M}$ to be a set of models in the *HFL*-structures, we could insist that $\mathscr{M}$ be some infinite sequence of models. For probabilistic representation, simply take some largest Boolean algebra of terms (more precisely, equivalence classes of terms *A, B* under the relation $\mathscr{L} \vdash A = B$) for which Fetzer and Nute's truth conditions give probabilities. But that is an additional constituent of *HFL*-structures which has to be accommodated in characterizing the *intended* model of the theory, for satisfaction of the condition of interpretation/idealization. Note that there is no "natural" ordering of the *worlds*, whereas the actual limiting frequency theory is able to take advantage of the natural temporal order of *events*. The difficulties elaborated above are indeed deep and sticky problems for the hypothetical limiting frequency interpretation in connection with the condition of interpretation/idealization.

I conclude that while the hypothetical limiting frequency interpretation is *superior to the actual limiting frequency view as far as conceptual adequacy is concerned*, it is *inferior with respect to interpretation/idealization*.

Long-Run Propensities

The theoretical advantage, discussed in the previous section, of the hypothetical limiting frequency conception over the actual limiting frequency conception was that the former deals with the possibilities that the history of the actual world is *finite* and that it *may exhibit coincidences that are not representative of the relevant physical features of the world*. Thus, the artificial restriction of the actual limiting frequency of tails, in tosses of honest coins when the moon, Mars and Halley's comet are in close opposition, to the values 0, $\frac{1}{3}$, $\frac{2}{3}$, and 1 results from the *finitude* of the relevant sequence; and the fact that the actual limiting frequency turned out to be, say, $\frac{1}{3}$, rather than, say, 0 or $\frac{2}{3}$, is a *coincidence that doesn't appropriately reflect the relevant physical facts*, e.g., the symmetry of the coin, the physical "honesty" of the toss, etc. By considering a hypothetical *infinite extension* of the actual sequence of such tosses, we do not artificially limit the possible values of the limiting frequency to the values $i/3$, for $i = 0, 1, 2, 3$. And considering *infinitely many such* hypothetical extensions of the actual sequence of just three tosses is supposed to accommodate the possibility that even in a lawful future world (hypothetically extended to include an infinite sequence of tosses under the relevant circumstances) the limiting frequency could be 0 or 1, or anything in between, however "improbable" such a value may be – where, as we have seen, characterizing the appropriate sense of "improbable" here is a sticky problem for hypothetical limiting frequentists in connection with interpretation/idealization.

But perhaps (at least part of) the motivation for adopting what has come to be called a *propensity* view of probability stems from difficulties that confront even hypothetical relative frequency views in connection with probabilities of single events. Let x_1, x_2, and x_3 be the three actual tossings of a fair coin under the celestial circumstances described earlier. And suppose that the actual relative frequency of tails in these three tosses is, in fact, $\frac{1}{3}$. We are interested in what the probability is that x_3 results in tails. Intuitively, let us assume for now, this single case probability is $\frac{1}{2}$. Let 'B' denote, in the model $M^* = \langle U^*, R^* \rangle$ corresponding to the actual world, the property of being an honest coin tossed honestly under the celestial circumstances described earlier, and 'A' the property of coming up tails. Should we identify the single case probability of x_3's coming up tails with the hypothetical limiting frequency theory's construal of $P(A/B)$? Suppose that, on this construal, $P(A/B)$ were $\frac{1}{2}$, as we should expect (given, of course, some adequate solution to the problem of interpretation/idealization for the hypothetical limiting frequency theory). This would be some evidence in favor of the hypothetical limiting frequency interpretation's applicability to single cases, as well as its appropriate applicability to sequences of events. But there are reasons why $P(A/B)$, under the hypothetical limiting frequency construal, may be equal to $\frac{1}{2}$ *other* than each element of a hypothetical infinite extension of $R^*(B)$ having, intuitively, a single case probability of $\frac{1}{2}$ of coming up tails. Suppose, for example, that during the first phase of any occurrence of the celestial configuration described – which lasts for half of the time of any such close opposition – all physically symmetrical coins are, in fact, *physically biased* roughly 2:1 for tails, while during the rest of the time of such a close opposition, all such coins are physically biased roughly 2:1 for heads, where the celestial configuration during these times is responsible, causally, for these biases. (Perhaps an example involving the tides would be more intuitive here.) Then $P(A/B)$, as a hypothetical relative frequency, should *still* be expected to be about $\frac{1}{2}$ – since about "half" (i.e., limiting relative frequency of $\frac{1}{2}$) of the elements of a hypothetical infinite extension of $R^*(B)$ may be expected to occur during the first phase, and the other "half" during the second phase, of close opposition – but, intuitively, if x_3 actually takes place during the second phase, the single case probability of x_3's being a member of A should be roughly $\frac{1}{3}$.

Some ways to accommodate such cases as this readily suggest themselves, the basic idea behind all of them being that the reference class must be chosen properly. In the above case, for example, the appropriate class is not an extension of

$R^*(B)$, but rather an extension of $R^*(B')$, the class of tosses of coins *during the second phase* of the celestial configuration described. But how shall the appropriate reference class be characterized in general?

One possibility is to say that the single case – x_3 in our example – should first be characterized uniquely by some set of permissible predicates – predicates permissible for the description of a reference class. Say that $F_1, \ldots, F_n$ are all permissible predicates and that x_3 is the only event in the course of the world's actual history that satisfies each of these predicates. Then we might identify the probability of x_3's being in A as the hypothetical limiting frequency of A in the class B' determined by $F_1, \ldots, F_n$. Unlike the actual limiting frequency view, it doesn't follow for hypothetical limiting frequencies that $P(A/B')$ is either 0 or 1. But this suggestion will not do, of course. And the reason is that $F_1, \ldots, F_n$ need not have anything to do, physically, with whether x_3 results in heads or tails, in order for them uniquely to pick out the event x_3 from all other events in the course of the world's history. If x_3 were the *only* actual event of a coin's being tossed under the celestial circumstances described above, then the F_i's need only to describe that celestial configuration, so that, as we have seen, the hypothetical limiting frequency of A in B' may be $\frac{1}{2}$, even though the correct single case probability of x_3's resulting in tails is, intuitively, $\frac{1}{3}$.

Two ways of accommodating this problem, consistent with a hypothetical limiting frequency conception of probability, suggest themselves. They are rough analogues for the hypothetical limiting frequency view of the two proposed solutions of the reference class problem for the actual limiting frequency interpretation, considered two sections back: (1) say that the appropriate class B' is the one determined by *every* permissible predicate actually satisfied by x_3, and (2) say that the appropriate class B' is the broadest ontically homogeneous reference class for x_3 and A, in the sense of the third section of this paper, i.e., the broadest class such that $x_3 \in B'$ and for any $B'' \subseteq B'$, $P(A/B'') = P(A/B')$, where P, here, is still hypothetical limiting frequency. Thus, according to (1), an appropriate hypothetical infinite extension of $\{x_3\}$ will be a "narrowest" class, while, according to suggestion (2), such an extension will be a "broadest" class, not all elements of which satisfy every permissible predicate which x_3 satisfies.

Suggestion (1) could be criticized on the ground that causally irrelevant factors should not be included in the description of the reference class. But Eells and Sober (1983) argue that the *values* of hypothetical limiting frequencies will not be affected by the specification of causally irrelevant factors. But there is still a difficulty with suggestion (1), in connection with probabilistic *explanation*. Salmon's (1971) well-known counterexamples to Hempel's requirement of maximal specificity tell against the suggestion. I shall not rehearse these considerations here in detail. But the basic idea is that if we must specify *all* permissible predicates in the reference class description, then causally and explanatorily irrelevant factors will be specified as well as those that are relevant. But it seems that in explaining why an event exhibited some attribute, we should assign the event to a reference class determined by only the causally or otherwise explanatorily relevant factors, to avoid citing explanatorily irrelevant factors in the explanation of the character of the event.

As to suggestion (2), a simple example of Fetzer's (1981, p. 91; also 1977, pp. 199–200) – also in an explanatory context – tells quite conclusively, it seems to me, against the idea. Suppose that Jones died of a brain tumor. Of course not everyone who has a brain tumor dies of it; assume, in fact, that brain tumors of the kind Jones had are irreducibly probabilistic causes of death. Say that $P(D/R \cap T)$, the hypothetical relative frequency of death among 60 year old human males – etc. – with a brain tumor, is r, in fact the *correct* "single case" value for Jone's death. But suppose it is also true that $P(D/R \cap H)$, the hypothetical limiting frequency of death among 60 year old human males – etc. – *with a certain serious kind of heart disease*, is *also* equal to r. Then $R \cap T$ is not a *broadest* ontically homogeneous reference class for Jones and D. Any such reference class must have $(R \cap T) \cup (R \cap H)$, i.e., $R \cap (T \cup H)$, as a subset. Suppose that $R \cap (T \cup H)$ in fact *is* a broadest ontically homogeneous reference class for Jones and D. Under the suggestion that single case probabilities are hypothetical limiting frequencies in broadest ontically homogeneous reference classes, the event of Jone's death cannot be probabilistically explained by assigning him to the class of 60 year old human males *with a brain tumor* (etc.) and citing the probability, r, of death in this class, but only by assigning him to the class of 60 year old human males (etc.) that *either* have a brain tumor *or* have that heart disease. The rationale behind taking the broadest homogeneous class was to *avoid* including causally *irrelevant* factors in explanation, but this formulation will *prohibit*, in cases such as this, specification of the *distinctively relevant* causally relevant factors.

Thus, suggestion (1) should be rejected because it *requires specification of causally irrelevant factors* in the description of the reference class, and suggestion (2) should be rejected because in some cases it will *prohibit specification of some factors that are distinctively causally relevant* for some single cases. These considerations, involving the conceptual adequacy of the hypothetical limiting frequency interpretation in connection with probabilities of single events, suggest an alternative conception of probability – a revision of the hypothetical limiting frequency view – on which the appropriate reference sequence is characterized not in terms of its subsequences and supersequences (and the relevant attribute and the single case in question), but rather in terms of the *operative causal conditions*, i.e., *the distinctively causally relevant factors*, themselves, where, after all, it was the failure of suggestions (1) and (2) to capture exactly these conditions that rendered them inadequate. Thus, Popper endorses the following alternative to suggestions (1) and (2):

> the frequency theorist is forced to introduce a modification of his theory – apparently a very slight one. He will now say that an admissible sequence of events (a reference sequence, a 'collective') must always be a sequence of repeated experiments. Or more generally, he will say that admissible sequences must be either virtual or actual sequences which are *characterized by a set of generating conditions* – by a set of conditions whose repeated realization produces the elements of the sequence. (1959, p. 34)

And Popper explains that, unlike frequency interpretations which take probability to be a property of sequences, somehow appropriately identified (or not), the propensity interpretation takes seriously the idea that an appropriate "sequence in its turn is defined by its set of *generating conditions*; and in such a way that probability may now be said to be *a property of the generating conditions*" (1959, p. 34). And Popper takes one more step. From the premise that actual and virtual frequencies depend on the experimental generating conditions, he concludes that "we have to visualize the conditions as endowed with a tendency, or disposition, or propensity, to produce sequences whose frequencies are equal to the probabilities; which is precisely what the propensity interpretation asserts" (1959, p. 35). As to probabilities of single events,

> now we can say that the singular event a possesses a probability $p(a, b)$ owing to the fact that it is an event produced, or selected, in accordance with the generating conditions b, rather than owing to the fact that it is a member of a sequence b. (1959, p. 34)[6]

Thus, evidently, where B^* is an "experimental arrangement", the propensity theory's interpretation of '$P(A/B^*) = r$' is (roughly): B^* possesses a *universal* (or "*almost universal*") *disposition* to produce, if repeated often, sequences B such that the limiting relative frequency of A's within B is r. (The reason for the qualification "or 'almost universal' " is the same as that encountered in the previous section, as discussed below.) Thus, this "long run"[7] propensity theory invokes just two concepts not present in the hypothetical limiting frequency theory investigated in the previous section: the idea of an *experimental arrangement* and the idea of a *certain kind of disposition of universal* (or "*almost universal*") *strength* with which some experimental arrangements are endowed. Let us now consider the effect of introducing these two new ideas on the *conceptual adequacy* of the theory and on the possibility of satisfying the condition of *interpretation/idealization*.

As to their effect on conceptual adequacy, Fetzer has compared the hypothetical limiting frequency theory with the propensity theory in relation to the way in which they may be invoked in accounting for certain frequency patterns that occur in the course of the actual world's history, in a passage worth quoting:

> the difference between them is describable as follows: the dispositional interpretation provides a theoretical basis for accounting for these patterns in terms of the *system's initial conditions*, insofar as the occurrence of actual frequencies is explained by reference to the dispositional tendencies that generate them; while [hypothetical limiting] frequency interpretations, by contrast, yield an empirical basis for accounting for these patterns in terms of the *pattern's ultimate configuration*, since the occurrence of actual frequencies is explained by reference to the hypothetical frequencies which control them. Consequently, the kind of explanation provided by a dispositional interpretation for the occurrence of actual frequencies during the course of the world's history is broadly *mechanistic* in character, while the kind of explanation afforded by these frequency constructions for those same occurrences is broadly *teleological* in character. To the extent to which the progress of science has been identified with a transition from teleological to mechanistic explanations, therefore, there even appear to be suitable inductive grounds for preferring the dispositional to the frequency approach. (1981, pp. 77–78)

(Actually, Fetzer is here comparing the *single case* propensity view – which will be considered in the next section – with the hypothetical limiting frequency view, but these considerations apply equally, of course, in the comparison under examination here, as he later points out (p. 107).)

I agree that the "mechanistic" character of the long run propensity view constitutes a conceptual advantage for this view over the "teleological" hypothetical limiting frequency view, especially in connection with the problem of assigning probabilities to single events. For, as we have seen in the rejection of suggestions (1) and (2), above, it seems that an appropriate reference class for a single case probability cannot be characterized in terms just of (i) the membership of the single event in question in the class, (ii) the relevant attribute, and (iii) how hypothetical limiting frequencies change when the class is narrowed or broadened. The long run *propensity* view, on the other hand, specifies that a single case should be referred to the reference sequence of events *which are produced by the same experimental arrangement*, the intent of which is to hold constant just the controlling causal factors present in the single case in question. And this would surely seem to be an appropriate sequence, for what else could be relevant to the physical probability that a single event will exemplify a given attribute than the physical circumstances under which the single event occurs? It seems clear that if a theory of probability that interprets probability in terms of sequences of events (together with other ideas) is to apply adequately to single events, then it must refer the single events to *such* reference classes, where the relevant physical circumstances under which the single event in question occurs are replicated in every element. So it seems that we should conclude, then, that *as far as conceptual adequacy goes, the long run propensity view is superior to the hypothetical limiting frequency theory*.

But what about satisfaction of the condition of interpretation/idealization, which is supposed to identify the conception in a precise manner and give the interpretation empirical (or cognitive, or, at least, "pre-probability-theoretical") significance? How are the two ideas of dispositions of universal (or "almost universal") strength and of experimental arrangements to be accommodated in a formal solution to the problem of interpretation/idealization?

I have two main points to make in connection with accommodating the idea of dispositions of universal or "almost universal" strength. First, it should be clear that the same sorts of considerations as those advanced in the previous section in connection with the hypothetical limiting frequency theory show that the long run propensity theory must also utilize some idea of "*almost* universal" strength of dispositions, rather than the idea of *strictly* universal strength, if the theory is to be able to accommodate the idea that trials in a sequence of events may be, intuitively, independent of each other. For if the trials are independent, then all sequences of results are "equipossible" and "equiprobable" (assuming that the relevant single case probabilities are supposed to be $\frac{1}{2}$). And even *without* independence and *without* the relevant single case probabilities all being equal to $\frac{1}{2}$, still it seems that *any* sequence of results – and hence *any* limiting relative frequency – should be granted to be *possible*, so that it would be incorrect to explicate probability in terms of a *strictly universal* disposition of an experimental arrangement to display its "characteristic" relative frequency. But we have already seen the serious difficulties involved in one way of trying to characterize the "almost universal" (or "almost every world") idea, and now it seems that these also confront the long run propensity theory in connection with the condition of interpretation/idealization.

Second, it seems that accommodating the dispositional idea (however the "almost universal" idea might be made clear) is closely connected with accommodating the idea of the experimental arrangement. For, in the first place, it is experimental arrangements that are supposed to be endowed with the relevant dispositions, and, in the second place, not every "arrangement" can be said to possess a disposition of the relevant kind, as will be presently seen. Thus, it seems natural to try to characterize the relevant kind of disposition in terms of the appropriate kind of physical arrangement.

Suppose I have a coin tossing device which has a knob on it controlling a pointer which can be set at any position between 0 and 1, inclusively.[8] If I set the pointer at position r, then the device will toss coins with a bias, intuitively, of $r : 1 - r$ in favor of tails, for all r between 0 and 1, inclusively. The internal mechanics of the device are not important. Now clearly, to say just that a coin is about to be tossed by this machine is not enough to specify an experimental arrangement, in a sense appropriate to the long run propensity interpretation of probability. Since such a specification of the arrangement does not include a specification of the setting of the control knob, the arrangement, so specified, does not possess an almost universal disposition to produce any particular limiting frequency of tails.

But now let us change the "initial conditions": the coin tossing device is put together with another device which rotates the pointer slowly back and forth at a constant speed from the 0-position to the 1-position to the 0-position, and so on. Now if the device is constructed in such a way that it tosses coins rapidly at constant short intervals, we can imagine that the combined device has an "almost universal" disposition to produce sequences of tosses with a "characteristic" limiting frequency of tails of $\frac{1}{2}$. But clearly again, even though the arrangement will "almost certainly" yield a "characteristic" limiting frequency, we have not specified an experimental arrangement in a sense appropriate for the long run propensity theory. For this theory is supposed to apply to single events, and, intuitively, when the pointer crosses the $\frac{2}{3}$-position, the probability of the toss' landing tails is $\frac{2}{3}$, and not the "characteristic" limiting frequency of $\frac{1}{2}$ produced by the device.

Thus, suppose we include in the description of the arrangement a position of the pointer. Have we now succeeded in specifying an experimental arrangement in a sense appropriate for the long run propensity theory? It seems unlikely, even if this specification again yields a "characteristic" limiting frequency. For just as the position of the pointer clearly and overtly indicates a single case or short run bias, there will no doubt also be *other* factors pertaining to conditions in and around the device which likewise introduce biases: the perhaps random movement of the air around the coin, the humidity, the tidal conditions, small earth tremors, and so on.

Note that for the combined device, the "characteristic" limiting frequency of tails depends on how the rate of rotation of the pointer varies as it sweeps across the dial: in the example above, this rate was assumed to be slow and constant. But if the pointer moved more slowly when it is in the interval $\left[0, \frac{1}{2}\right]$ than when it is in the interval $\left[\frac{1}{2}, 1\right]$, then the "characteristic" relative frequency of tails would be less than $\frac{1}{2}$. If some proposed specification of the experimental arrangement does not specify the position of the pointer, then we may say that the possible positions of the pointer are unspecified possible initial conditions, and that the "characteristic" limiting frequency of tails depends on both the experimental arrangement, as specified, *and the "distribution of initial conditions"* (as Sklar (1970) expresses the idea). In the example above, this distribution is determined by a function $v(r)$ = the absolute rate of speed of the pointer across the point r on the dial, this assumed to be small and constant for all sweeps across r.

So it seems that there are two options open to the long run propensitist in connection with the nature of experimental arrangements: (i) specification of an experimental arrangement must include a specification of the distribution of unspecified initial conditions (as well as a specification of certain of the initial conditions), and (ii) all of the initial conditions must be held fixed. As to (i), three difficult problems arise. First, how is it to be decided *which* of the initial conditions are to be held fixed and which of them should be only partially specified by giving their distribution? Second, on what grounds should one distribution of the unfixed initial conditions to be preferred to another? Sklar has urged that "what this distribution would be [if the experiment were repeated often] is completely unconstrained by any *lawlike* features of the actual world whatsoever!" (1970, p. 363). On the other hand, Settle has reported private communication from Popper in which the latter conjectures that (in Settle's words) "there is a law of nature, that unless they are constrained, initial conditions have a ('natural') propensity to scatter over the interval left open to them by the (constraining) experimental conditions" (Settle 1975, p. 391). Now if the initial conditions *did* have a propensity to scatter over some interval with some characteristic distribution as a matter of law, then perhaps we would have an important improvement over the hypothetical limiting frequency theory: namely, a principle governing the extension of an actual sequence $R^*(B)$ to a sequence $R(B)$ in a B-maximal extension $M = \langle U, R \rangle$ of a lawlike future world. The principle would state that the initial conditions must be distributed over the elements of $R(B)$ according to how, as a matter of actual law, they must be distributed. But now we must ask whether this propensity to scatter over an interval with some "characteristic" distribution is a universal or an "almost universal" disposition, and if it turns out to be an "almost universal" disposition, then by now familiar problems emerge again. In any case, whether the disposition must be universal or only "almost universal" would seem to depend on a more precise formulation of the conjecture (e.g., are the different configurations of initial conditions *independent* of each other?), and perhaps on empirical investigation.

But, perhaps more importantly, the third difficulty with (i) pertains to the desire to make probability applicable to single events. If the exact configuration of initial conditions in a given trial (i.e., the actually obtaining values of the "hidden" variables) makes a physical difference with respect to the result of the trial, then it would seem inappropriate to leave any of the initial conditions unspecified.

This suggests consideration of alternative (ii). But the problem with suggestion (ii) is that it may very well be, as a matter of fact, a matter of physical law that in some cases, configurations of initial conditions cannot remain fixed from trial to trial. So if we have to consider a virtual sequence $R(B)$ in which all of the initial conditions remain unchanged from element to element, we may have to consider *nonlawlike* B-maximal extensions of lawlike future worlds. Also, if conditions such as being the nth element of $R(B)$ are to count as initial conditions, then there would also be logical difficulties with the idea of replicating an experiment, holding all of the initial conditions fixed. The single case propensity interpretation, to be considered in the next section, has, I think, a more plausible suggestion to offer along basically the same lines. So I shall postpone consideration of the "hold everything fixed" idea until then.

But perhaps it will be urged that we have been going about the explication of the two new concepts of the long run propensity interpretation in the wrong direction: instead of trying to characterize the ("almost") universal disposition idea in terms of the experimental arrangement, we should first try to characterize the disposition, and then, in terms of this, characterize the relevant kind of experimental arrangement. Of course a particular, individual experiment cannot itself be repeated, literally speaking, so what is needed, of course, is a characterization of the relevant kind of *experiment type*. And according to this new approach, two particular experiments will be of the same type (of the appropriate kind of type) if they are *both endowed with the same universal or almost universal disposition* (of the appropriate kind). If this idea could be worked out satisfactorily – in connection with formal adequacy and interpretation/idealization – then, again, the long run propensity theory would have important advantages over the hypothetical limiting frequency theory as considered in the previous section. For then, the long run propensity theory would be in possession of a principle governing the extension of actual sequences $R^*(B)$ to infinite virtual, or hypothetical, sequences $R(B)$ in B-maximal extensions of lawful future worlds: we may say that an *admissible* such sequence for the purpose of assessing a probability $P(A/B)$ would be one whose every member was endowed with the very same disposition (of the appropriate kind) with which every member of $R^*(B)$ – which may consist of just one trial – is endowed. Thus, perhaps the characteristic structures of the long run propensity interpretation would be like those of the hypothetical limiting frequency interpretation except for having an additional component: say a function D from pairs $\langle M, x\rangle$ into properties of events, where, of course, we should not insist that $D(M, x) \in R$, where $M = \langle U, R\rangle$. Then, in the intended model, for any world M and event x, $D(M, x)$ is the universal or almost universal disposition of the relevant kind with which $M(x)$ is endowed, and an *admissible* extension of a sequence $R'(B)$ consisting, say, of just one event $M'(x)$ – where $M' = \langle U', R'\rangle$ is a lawful future world – would be a sequence $R(B)$ in a world $M = \langle U, R\rangle$, every member of which has the property $D(M, x) = D(M', x) = D(M^*, x)$. This, plausibly, might accommodate also the conceptual difficulty with the hypothetical limiting frequency view that, intuitively, lawful future worlds, and their B-maximal extensions, may have *statistical* laws differing from those that hold in the actual world, though, by the definition of lawful worlds, the *universal* laws that hold of them are the same as those that hold in the actual world.

Of course the above considerations only constitute a step in the direction of *conceptual* and *formal* adequacy of the long run propensity interpretation, leaving the problem of *interpretation/idealization* untouched. A solution to the latter problem is supposed to identify the intended model and associate independently understood concepts, objects or phenomena with the constituents of the intended model. Of course counterparts of the problems encountered earlier in connection with the hypothetical limiting frequency theory for interpretation/idealization remain (e.g., what does "*almost* every world" mean?), but a new part of the problem for the long run propensity view is to interpret the new function symbol 'D' – in other words, to explicate the relevant kind of "almost" universal disposition to produce sequences with a characteristic limiting frequency. From an antagonistic point of view, this property may be characterized as whatever it is that every member of hypothetical infinite sequence must have in order for the limiting frequency of the relevant attribute in that sequence to be appropriately transferable to any member of the sequence. From the other point of view, the postulation of the existence of this thing has been characterized as "*a new physical hypothesis* (or perhaps a metaphysical hypothesis) analogous to the hypothesis of Newtonian forces" (Popper 1959, p. 38). And, indeed, what seems to be lacking in the long run propensity interpretation is an explication of 'D' in terms of *old* concepts, objects or phenomena that are *already* understood *independently of probabilities or propensities*. Below I shall consider the question of whether propensity theories should be required to satisfy a condition of interpretation/idealization (for 'D'), in light of the idea that such theories may be characterized as involving, after all, a *new* physical hypothesis to the effect that there are propensities, of some sort, which may be of a "new metaphysical category". But, independently of the appropriateness of the condition, it seems appropriate to conclude that, while for reasons given several pages back, the long run propensity interpretation is *superior to the hypothetical limiting frequency theory as far as conceptual adequacy goes*, it is, for lack of an appropriate interpretation of *'D', inferior with respect to interpretation/idealization*.

Singe-Case Propensities

There is a formulation of the single case propensity theory of probability that initially seems to have the advantage over the long run approach, in connection with interpretation/idealization, that the conception of probability offered can be explicated in terms understood independently of the ideas of universal or "almost universal" dispositions – i.e., independently of the intended interpretation of the function symbol 'D' of the intended model of the long run account, as described at the end of the previous section. The single case theory which I shall consider below is, essentially, that of Fetzer and Nute (1979, 1980; see also Fetzer 1981, pp. 49–73).

The basic idea is that instead of looking at relative frequencies in sequences of *repetitions of experiments in single worlds* (and then perhaps at the relative frequency of worlds' exhibiting a given frequency), we look first at *sequences of lawful future worlds in which the single event in question takes place*, exhibiting or not exhibiting the relevant attribute. Thus, suppose we wish to give truth conditions for the statement '$P(A/x) = r$', where x is the single event in question, say a toss of a coin, where A is the relevant attribute, say coming up tails, and where the statement means, intuitively, that the single case probability (propensity) of that toss' resulting in tails up is r. Let $M_1, M_2, \ldots$ be an infinite sequence of lawful future worlds, i.e., a sequence of worlds each of which obeys all the universal laws which the actual world M^* obeys and whose histories are the same as that of M^* at least up to the time of the event x. Then we may give truth conditions as follows:

$$\begin{aligned}&\text{`}P(A/x) = r\text{' is true in } M^* \text{ if and only if}\\ &\lim_{k\to\infty} \frac{\#\{M_i : i \leqslant k \text{ and `}Ax\text{' is true in } M_i\}}{k}\\ &= r.\end{aligned}$$

These truth conditions have the advantage over both the hypothetical limiting frequency interpretation and the long run propensity interpretation that they avoid the necessity of providing principles governing the extension of actual sequences of events to longer, ideally infinite, sequences of events that occur in some possible world. In particular, no recourse to the D-component of the long run propensity theory is necessary. Also, of course, there is no longer the problem of the reference class or experiment type, and, envisioning x as being, in every relevant world, numerically the very *same* event, we have, here, an interpretation that is truly applicable to the single case.

But there are difficulties with this approach that are similar to the problems encountered in connection with the hypothetical limiting frequency theory. Since there are, presumably, *nondenumerably* many lawful future worlds relevant to the probability statement in question (as argued in the fourth section of this paper), the problem arises of how to select from all of these worlds an appropriate *denumerable* sequence, $M_1, M_2, \ldots$, in terms of which the truth conditions for '$P(A/x) = r$' should be given. Also, as long as Ax and $\bar{A}x$ are both physically possible, it would seem that there should be infinitely many lawful future Ax-worlds as well as infinitely many such $\bar{A}x$-worlds, so that, for any value of r between 0 and 1, inclusive, there will be *some* sequence of worlds which, together with the suggested truth conditions, will yield the truth of '$P(A/x) = r$'. Thus again, Fetzer and Nute require only that "*almost every*" sequence of lawful future worlds satisfy the above truth conditions in order for '$P(A/x) = r$' to be true, where this is made precise as follows. Let $M^1, M^2, \ldots$ be an infinite sequence of infinite sequences of lawful future worlds, where M_i^j is the ith member of the jth sequence. Then revised truth conditions are suggested as follows:

$$\begin{aligned}&\text{`}P(A/x) = r\text{' is true in } M^* \text{ if and only if}\\ &\lim_{m\to\infty} \frac{\#\left\{M^j : j \leqslant m \text{ and } \lim_{k\to\infty} \frac{\#\left\{M_i^j : i \leqslant k \text{ and `}Ax\text{' is true in } M_i^j\right\}}{k} = r\right\}}{m}\\ &= 1.\end{aligned}$$

But the same sort of problem arises again. As noted above, if there are infinitely many Ax-worlds that are lawful and future, as well as infinitely many such $\bar{A}x$-worlds, then there is, for any value of r, some sequence of worlds which, on the first suggested truth conditions, yields the truth of '$P(A/x) = r$'. But if there is even one such sequence, there are infinitely many such sequences: simply reorder the first n terms, for each finite n, to get infinitely many such sequences. Arrange these sequences into a sequence of them, in any order, and you get a sequence of sequences of lawful future worlds which, together with the second suggested truth conditions, yields the truth of '$P(A/x) = r$'. And this can be done for any value of r, as long as there are infinitely many lawful future Ax-worlds as well as infinitely many such $\bar{A}x$-worlds. Thus, the problem is to specify principles for selecting an appropriate sequence of sequences of worlds from presumably nondenumerably many such sequences. We may take Fetzer and Nute's idea one step further and say that '$P(A/x) = r$' is true if and only if "almost every" sequence of sequences of lawful future worlds satisfies the second suggestion, where *this* could be made precise in terms of an infinite sequence of infinite sequences of infinite sequences of lawful future worlds, but it is obvious that the same sort of problem would arise again. And so on.

Perhaps there may be another way of rigorously capturing the intuitive idea that the proportion of Ax-worlds in a random selection from all lawful future worlds will almost certainly be about r. Indeed, the difficulties with the above approach may suggest a measure theoretic approach, according to which '$P(A/x) = r$' is supposed to be true if and only if $r = \sum_M \textit{Probability}(M)$, where the summation is taken over lawful future Ax-worlds M, or, more generally, $P(A/x) = \int_{|Ax|} \mu(M) d\mu(M)$, where, again, the integral is over lawful future worlds in $|Ax|$. But now it is natural to ask where the probability function *Probability* comes from, or where the density function μ on worlds comes from.

Suppes (1973) has formulated a measure theoretical approach in which a probability function *Probability* can be inferred from an axiomatized quaternary relation $A/B \geq C/D$ (meaning that the propensity of A's occurring given an occurrence of B is at least as great as the propensity of C's occurring given an occurrence of D). But then the problem remains of giving a *physical interpretation* of the relation $\geq$, in the sense elaborated earlier in this paper. Note that this is different from the problem of *application* in particular physical contexts, whose solution may simply require an association of kinds of physical events with *A, B*, etc., and the assumption of additional axioms appropriate to the particular physical context. Suppes (1973) gives an example of this, relating to the phenomenon of radioactive decay.

Also, Giere (1976) formulates a measure theoretical kind of approach, in which finite, countable and continuous *possibility structures* are defined. These involve, basically, a set of possible worlds, a partition of this set, and (i) in the case of finite possibility structures, an equal measure over the possible worlds, (ii) in the case of countable possibility structures, an "equal measure" on the possible worlds obtained from a one-to-one correlation between the possible worlds and the interval [0, 1] and a uniform density on [0, 1], and (iii) in the case of continuous possibility structures, something more complicated. In each case, a probability space on the partition of worlds can be inferred, where this is used to give a formal definition of a "propensity function" on the set of final states of a stochastic system, where this set is an isomorph of the partition of worlds. In each case, the basic idea is, as Giere puts it, that "physical probabilities are a measure of the density of possibilities open to a system in a given initial state" (p. 338), where these possibilities correspond, in the formal theory, to the " 'equipossible worlds' " (p. 338). But problems are: How to characterize possible worlds in such a way that they *are* "equipossible" (Bertrand paradoxes)? What does "equipossibility" mean? And how to come up with an appropriate "density of possibilities" (density function) in any given case? Now again, I do not think that all of these problems will be serious in all contexts of *application*, as it is clear from the examples that Giere gives that his formulation can be accommodated to many different kinds of stochastic phenomena, including radioactive decay and Bernoullian sequences. But this just displays the *formal adequacy* of the theory as applied to various kinds of phenomena, where the question under consideration is: How are the entities of the abstract possibility structures to be related to something independently understood in order that we may have an explication of propensity *via* interpretation/idealization?

In connection with his formulation of the propensity theory in the case of finite possibility structures, Giere himself states (where it is clear that these comments apply also to the countable and continuous cases),

> The individual worlds and the uniform measure need have no direct physical correlates. That is the respect in which this semantics is merely a formal semantics. It would be an interesting physical hypothesis that underlying every stochastic process there exists a set of physically equipossible states. The above account of physical propensities is compatible with this hypothesis but does not require it. In any case, I doubt it is true.
>
> Metaphysically speaking, then, what are physical propensities? They are weights over physically possible final states of stochastic trials – weights that generate a probability distribution over sets of states. The [uniform measure] function u provides only a formal and rather shallow analysis of this distribution. But is it not just the task of a propensity interpretation to explain what these weights are? No, because it cannot be done. We are faced with a new metaphysical category. (1976, p. 332)

And he suggests that what he has in mind may be the same point as Popper (1959, p. 38) expressed by saying that his propensity theory involved a "*new physical hypothesis* (or perhaps metaphysical hypothesis) analogous to the hypothesis of Newtonian forces".

This feature of single case propensity theories is also found in Fetzer's (1981, p. 41) theory, and he takes the idea a step further with a dispositional ontology according to which, for example, "*individual objects* are continuous sequences of instantiations of particular arrangements of [universal and statistical] dispositions" and "*singular events* are continuous sequences of instantiations of particular arrangements of [universal and statistical] dispositions" (1981, p. 42). Here, dispositions are ontologically primitive, but

> a dispositional predicate ... may be informally defined as a set of ordered triples, each of which consists of a test trial description T^i, an outcome response description O^j, and a numerical strength specification r^k, i.e., $\{\langle \mathrm{T}^1, \mathrm{O}^1, \mathrm{r}^1\rangle, \langle \mathrm{T}^2, \mathrm{O}^2, \mathrm{r}^2\rangle, \ldots\}$, where the number of members of these sets is determined by the variety of different trial tests and response outcomes that are ontological constituents of each specific disposition – a possibly infinite set. (1981, p. 37)

Thus it is clear that Fetzer, too, proposes the existence of propensities as "a new physical hypothesis", where the single case propensity semantics with which this section began, involving infinite sequences of infinite sequences (and so on?) of trials under identical conditions are "required to display the complex character of propensity attributions" (private communication), but are not *definitive* of single case propensities. Whatever the detailed relation is between the propensities themselves and the semantics proposed, it is clear that for Fetzer as well, single case propensities are not *definable* in terms of independently understood concepts and phenomena.

That single case propensities are not definable in terms of independently understood concepts and phenomena (with the exception, in some cases perhaps, of a qualitative propensity relation) seems inevitable, if they belong to a "new metaphysical category". Nevertheless, it might be hoped that empirical significance could be given to the concept of single case propensities by describing procedures whereby hypotheses involving the concept may be tested, hypotheses such as, "The single case propensity of *this* nucleus' decaying before the end of 30 years is $\frac{1}{2}$"; and perhaps then, in terms of procedures for testing *such* hypotheses, we may be in a position to test the physical hypothesis that single case propensities obey some axiomatization of probability.

But the following problem naturally arises in connection with the idea of testing single case propensity statements: since, in any given single case x, the relevant attribute, say A, either occurs or fails to occur, the difference between, for example, $P(A/x) = 0.7$ and $P(A/x) = 0.8$ would seem to make no difference in experience. (See Reichenbach (1949, pp. 370–371) for another statement of this argument.) Despite the unrepeatability in principle of particular single events, however, a number of propensitists have suggested that single case propensity statements may be tested by reference to certain relative frequencies. Thus Fetzer says,

> although a single case probability statement surely does pertain to singular events, it is somewhat misleading to suppose that the events involved only occur a single time. A single case propensity statement, after all, pertains to every event of a certain kind, namely: to all those events characterized by a particular set of conditions. Such a statement asserts of all such events that they possess a specified propensity to generate a certain kind of outcome. Consequently, although the interpretation itself refers to a property of conditions possessed by single events, such statements may actually be tested by reference to the class of all such singular occurrences – which surely is not restricted to a single such event. (1971, p. 478)

But this would seem to raise the old problem of the reference class, or of the description of the experimental arrangement: *what is* the "certain kind" of event to which we should think of the single case in question as belonging, or *what is* the relevant "particular set of conditions"? Note that here, the reference class problem (or experiment type problem) arises in connection with the problem of *testing* probability statements, while for the relative frequency view and (as presented above) for the long run propensity view, this was a problem for the very explication of probability. And I would agree that a problem of this kind need not be as serious for a theory of probability if it arises *only* in connection with its theory of testing and *not* for the very explication of probability. But recall that we have decided to look for a kind of explication of the idea of single case propensities (for empirical interpretation) in empirical methods of testing propensity statements. So perhaps the problem *is* serious, as the following considerations may further suggest.

Now Fetzer (1981, pp. 50–51) has given an answer to the question posed above just after the quotation. He advances a requirement of maximal specificity according to which, roughly adapted to the context of this discussion, the "particular set of conditions" must include all "nomically relevant" factors, where a factor is nomically relevant, roughly, if its presence or absence in a given single case event would affect the single case propensity of that event's having the relevant attribute.[9] Thus, it would seem that in order to test a single case propensity statement by looking at a relative frequency generated by an appropriate "particular set of conditions", we must either already know or have good reason to believe or also test *other* single case propensity statements, namely, statements to the effect that (i) some particular set of conditions is present in every trial and (ii) that set of conditions is appropriate in the sense given above (all nomically relevant factors are held fixed in every trial) so that we may know that (iii) the single case propensity for the relevant attribute is the same in each trial. (Cf. Fetzer 1981, pp. 248–254.)

Giere (1973, p. 478) has made the same point about testing single case propensity statements, in connection with tests of hypotheses about the half-lives of radioactive nuclei. He describes the standard procedure, which "assumes that each nucleus in the sample has the *same* half-life, whatever its value. Thus the test assumes the truth of some propensity statements, though not of course the truth of the hypothesis being tested". But he goes on to argue that this feature of testing single case propensity hypotheses is not unique in science:

> Consider the concept of an individual (as opposed to total) force in classical physics. Any attempt to determine the value of a particular force requires assumptions concerning other forces, e.g., that there are none operating or that their influence has been taken into account. Thus, if one regards the concept of an individual force as a legitimate empirical concept, one cannot dismiss single-case propensities solely on the ground that empirical tests of propensity hypotheses assume the truth of other propensity statements. (p. 479)

But it seems to me that the two cases are not parallel and that one can measure individual forces under significantly weaker assumptions than one can measure single case propensities. Consider the single case propensity hypothesis, "The probability that *this* nucleus will decay within 30 years is $\frac{1}{2}$." As Giere explains, the standard procedure for testing such a hypothesis is to obtain a large number of nuclei of the same kind and then count the numbers of them that decay within specified periods of time. The single case propensity hypothesis that such a test assumes to be true is that all the nuclei in the sample have the *same* half-life, whatever its value. How does one test a hypothesis concerning an individual force, say the force exerted by a certain spring when its length is two inches? One may first measure the total force present in the absence of the spring (say by observing the acceleration of some object) and then measure the total force present when the spring is introduced (say by observing the acceleration of an object when placed at the end of the spring in its two-inch configuration), and then calculate the difference between the two values. Presumably, the hypothesis concerning forces that Giere would say is assumed in such a test is that, when the spring is introduced, all of the *other* individual forces remain the same.

But note that the case of individual forces is quite different from that of single case propensities in that the individual forces can each be *separately measured* by *separate* tests which do not require the *same* assumptions in each case. In principle at least, each individual force can be eliminated and the total remaining force can be compared to the original force to yield a measure of that force. In each case, of course, one assumes that all the remaining forces do not change, but since for each individual force, the remaining individual forces will constitute a *different* set of individual forces, the assumptions made in the different cases will not all be the same. Also, the law of addition of forces can in principle be tested by determining whether or not the sum of all the values determined for the individual forces add up to the value determined for the total force. (In practice, of course, this can only be done in the context of a background configuration of forces that cannot in practice be eliminated, but of course one can obtain additional confirmation for a hypothesis concerning an individual force by looking at the effect of introducing the relevant conditions in many different constellations of background forces.) In the case of the single case propensities of the individual radioactive nuclei decaying within 30 years, one cannot test each nucleus (of the same kind) separately: for each nucleus, the test of the relevant hypothesis is the very same test. Another difference is that in the case of individual forces, no assumption whatsoever must be made about the value of the individual force in question, whereas in the case of single case propensities, one must assume that the nucleus in question has the same propensity to decay within 30 years as do all the other nuclei in the sample.

Now it may be insisted that the analogy which Giere urges still holds, for in each case, it is still necessary to make *some* assumptions concerning the relevant kind of individual thing (propensity or force). But in the case of force, the method of testing itself makes it clear that *it is an individual kind of thing that is being measured: individual* forces, indirectly by measuring *different* pairs of *different* total forces and calculating differences. But in the case of testing single case propensity hypotheses, the method of testing does *not* give empirical significance to the idea that it is an *individual* or *single case* propensity – rather than, for example, a *long run* propensity – that is being measured. There is nothing in the general sketch of the method of *testing* single case propensity hypotheses under consideration which distinguishes the single case propensity interpretation from the long run propensity view.

Here is another aspect of this kind of difficulty in testing distinctively *single case* propensity hypotheses. That single case propensities exist is a new physical hypothesis. And until and unless this hypothesis is developed in more detail to the contrary, it would seem that two objects or sets of experimental conditions may differ in no respects whatsoever except for a certain single case propensity with which they are endowed (and, of course, the physical consequences, e.g., pertaining to relative frequencies generated, of having the different single case propensities). On single case propensitist principles, it would seem conceivable, for example, that there is a certain "kind" of radioactive nucleus whose half-life has been tested extensively, where the *only* difference between individual nuclei of this kind is that they can actually have *different* ("single case") half-lives ("statistical hidden variables"), where the half-lives of these nuclei (which vary considerably) are distributed among such individual nuclei in such a way that testing random samples of them in the standard way has always yielded the same result as would be expected if they all had the *same* ("single case") half-life (which, say, scientists have been assuming): the different half-lives are distributed homogeneously among the nuclei of the relevant "kind". Then how to test the *single case* propensity hypotheses pertaining to the different nuclei? And how to test whether all the nuclei have the *same* half-life, or *different* half-lives so distributed among the nuclei that relative frequency tests give the same results as they would if all the nuclei had the same half-life? Now these considerations may only be valid in the absence of a well-developed theoretical background, here pertaining to how the structure of a nucleus (which may determine a *kind* of nucleus) is related to its half-life. But if Giere (1973, p. 478) is correct in suggesting that "in the absence of a well-developed theoretical background, observed relative frequencies may provide the only evidence for propensity statements", then examples such as this one strongly suggest that the single case and long run propensity conceptions of probability cannot be distinguished in terms of empirical methods of testing the relevant propensity hypotheses.

At the beginning of this section, we considered a formulation of the single case propensity theory of probability that initially seemed to be superior to the long run propensity theory with respect to interpretation/idealization. But problems for that formulation arose, problems of the same general kind as arose earlier for the hypothetical limiting frequency interpretation and which also confront the long run propensity theory. Recall that the long run propensity theory's solution to the problem of actual sequence extension was – in terms of the formulation I suggested in the fifth section of this paper – to introduce a new two-place function D, which, in the intended model, has as its range of values a set of dispositional properties of a certain kind. While the introduction of D rendered the long run propensity view superior to the hypothetical limiting frequency theory with respect to the condition of conceptual adequacy, it also rendered it inferior with respect to interpretation/idealization. Now it seems that the single case propensity view must do something very similar: it too must introduce a two-place function, say D^*, whose range, in the intended model, would be a set of dispositions, this time "single case statistical dispositions", rather than universal (or "almost universal") dispositions to produce statistical displays in the form of characteristic relative frequencies. Thus, temporarily leaving aside the question of conceptual adequacy and of the relevance of the condition of interpretation/idealization to propensity theories, it seems that the single case interpretation is *at least* as bad off as far as interpretation/idealization goes as is the long run view.

Just above, a third possible way of securing interpretation/idealization was considered – where the first two ways, of course, were *via* possible worlds and sequences of various kinds and *via* all this plus D, or D^*. The third was in terms of empirical procedures for testing hypotheses of the relevant kind, which, if successful, should help the theory to obtain at least empirical *interpretation*. But it seems that this approach cannot distinguish between the long run and single case dispositional approaches: both approaches would use the available finite relative frequencies in the very same way in tests of the relevant hypotheses. Roughly and intuitively speaking, the two theories *idealize* the relevant phenomena in different ways (one in terms of D and the other in terms of D^*), where the objects of *interpretation* of the two idealizations are the same: observable finite relative frequencies. Still leaving aside the question of conceptual adequacy and of the relevance of the condition of interpretation/idealization to propensity interpretations of probability, we can ask which of the two theories under consideration is better off in connection with the interpretation/idealization condition by asking: Which of the two idealizations is more fully interpreted in terms of the available relative frequencies? That is, we ask, intuitively: Which theory commits itself to the stronger concept (idealization), the concept a higher proportion of the content of which will therefore *lack* interpretation in terms of the common objects of interpretation? And it seems clear that it is the single case conception which is the stronger concept. The long run concept is, roughly, a two-component concept: the concept of universal (or "almost universal") dispositions *plus* the idea of limiting frequencies, where the second component of the concept is quite well understood. The single case concept, however, is a kind of *one-component* "organic" union of the two components of the long run concept: the concept of a *partial or statistical disposition of a specific strength*, rather than a *universal* (or "*almost universal*" disposition to produce a statistical *display*. And in this case it seems that the whole (the concept of a *statistical disposition*) is greater, or stronger, than the sum of its parts (the concept of a *universal or "almost universal" disposition* plus the concept of a *display which is of a statistical character*, i.e., a sequence with a characteristic relative frequency). The single case propensity concept compresses the idea of a display of a statistical character into the concept of the disposition itself.

The question for interpretation/idealization is, here, how adequate the interpretation is relative to the concept, or intended idealization – that is, in this case, the extent to which observed relative frequencies capture the features or components of the idealization, or proposed concept. For the long run theory, what we always (or "almost always") observe (or, at least what we can in principle observe) is the *direct manifestation* of the relevant disposition: the disposition is a disposition to produce sequences of events with a characteristic limiting frequency, and we can, in principle, observe a sequence with a characteristic frequency of the relevant attribute. What we don't actually observe, of course, is the disposition itself. Thus, what we can in principle observe is the physical interpretation of one component of the two-component long run concept, and this is the direct manifestation of the relevant disposition. For the single case propensity view, on the other hand, what we actually can observe (relative frequencies in sequences) is *not* the direct manifestation of the relevant disposition, it seems. For the statistical disposition is supposed to operate directly on the single case, and *via its direct operation on single cases* it controls the observed relative frequencies in accordance with Bernoulli's theorem. Thus, for the single case propensity interpretation, there are *two things which we do not observe when we observe relative frequencies: namely, the direct* manifestation of the disposition, as well as, of course, the disposition itself.[10] Since (i) *observed relative* *frequencies* (that which it seems that both theories must use to secure, empirically, interpretation/idealization) are the *direct manifestations* of the relevant disposition on the long run view and also the *physical interpretation* of the other component of the two-component long run concept, and since (ii) they are *not* the direct manifestations of the single case disposition, I conclude that, *in relation to the condition of interpretation/idealization, the long run propensity view fares better than the single case interpretation*.

Let us now turn to the comparison between the single case and long run propensity interpretations with respect to the condition of conceptual adequacy. We have already seen that both theories are, as Fetzer puts it, "broadly mechanistic" in character rather than "broadly teleological", like actual and hypothetical limiting frequency theories. But what of some of the other desiderata that should be brought to bear? Consider the problem of attributing probabilities to single cases, say the probability that event x will exemplify attribute A. Suppose that, as the long run propensitist requires, there really is a dispositional property $D(M^*,\ x)$; and suppose that, as the single case propensitist requires, there really is a single case propensity $D^*(M^*,\ x)$ for x itself to exhibit attribute A. (What these suppositions amount to, as far as a detailed *explication* of the concepts is concerned, is a problem for interpretation/idealization; here, we are interested only in the theoretical consequences of what is *intended* by advancing the concepts.) Now the possession of the property $D(M^*,\ x)$ by each member of a hypothetical infinite sequence of events in a lawlike future world is supposed to guarantee (or "almost guarantee") that the limiting relative frequency of A in the sequence is, say, r. But, as single case propensitists emphasize, what holds in the long run does not always matter in the single case (Hacking 1965, p. 50; see also Fetzer 1981, pp. 110–111).

Although Hacking's example (see the reference) pertains to *rational decision making*, the following considerations are intended to show that, as far as *physical probabilities* are concerned, it is also true that what matters in the single case need not matter in the long run (on the conceptions of single case and long run propensities under consideration). The possession of $D(M^*,\ x)$ by every member of a hypothetical infinite sequence of events in a lawful future world would seem to be compatible with (at least, is *not obviously incompatible* with) the members' possessing single case propensities for exhibiting A that differ from case to case, and differ from r. If both $D(M^*,\ x)$ and single case propensities exist, then all the possession of $D(M^*,\ x)$ by the members of a hypothetical infinite sequence has to guarantee (or "almost guarantee") in order to guarantee (or "almost guarantee") that the limiting relative frequency of A is r, is merely that the *average* of the single case propensities for A in the sequence is r, where this idea of an "average" can be made precise in the obvious way. And indeed, on Popper's conception, it *would* seem possible for possession of $D(M^*, x)$ by every member of a hypothetical infinite sequence in a lawful future world to guarantee *merely* that the *average* of the single case propensities is r, given his conjecture about configurations of initial conditions distributing themselves over the interval left open to them *as a matter of physical law*, and, hence, in lawlike future worlds (see the fifth section of this paper above on this conjecture, and Settle 1975, p. 391). And then it would seem that even though *different* configurations of initial conditions would give rise to *different single case propensities*, still the *lawlike distribution* of the configurations of initial conditions over the interval left open to them would guarantee (or "almost guarantee") the *characteristic limiting frequency*. Thus, what matters in the single case need not matter in the long run, assuming the truth of Popper's conjecture. Until the concept of $D(M^*,\ x)$ is refined in such a way that it can be shown that possession of $D(M^*,\ x)$ by the relevant events cannot guarantee merely an average single case propensity for the relevant attribute among the single events in question – where the refinement does not make possession of $D(M^*,\ x)$ conceptually equivalent to the possession of a single case propensity – it would seem that as far as attributing probabilities to single events is concerned, the single case theory is conceptually superior to the long run interpretation.

Having considered the single case adequacy of the long run propensity approach, what now about the long run adequacy of the single case approach? According to Fetzer,

> The most important benefit of the "single case" approach . . . is that it not only accounts for the meaning of *single case* probabilities but also solves the problem of *long run* probabilities; for, given the values of the relevant single case probabilities, calculations of long run probabilities for the various combinations of outcomes over various lengths of trials may be made on the basis of the mathematical principles [such as Bernoulli's theorem] for statistical probabilities. Thus, the fundamental advantage of the single case interpretation is that it yields a construct which is theoretically significant for *both* the long run *and* the single case . . .. (1981, p. 111)

Aside from whatever may be said in favor of the idea that some probabilistic phenomena are not "grounded from below" in terms of probabilistic laws on the level of individuals, but are rather "imposed from above" (see Hacking 1980, and Baird and Otte 1982 on this), this, of course, is correct; where, however, if some probabilistic phenomena actually were "imposed from above", then perhaps in such cases a long run approach would be more appropriate. But, aside from such worries, and given the general promise of Poisson's law of large numbers program of grounding probabilities from below, and given the theoretical difficulties of the long run conception in the single case, it seems appropriate to conclude that – though the long run propensity view may be superior to the single case theory in connection with interpretation/idealization – as far as conceptual adequacy is concerned, the single case propensity interpretation fares better than the long run theory.

"New Metaphysical Category"

I have argued that to the extent to which philosophical theories of objective probability have offered theoretically adequate *conceptions* of objective probability, in connection with such desiderata as causal and explanatory significance, applicability to single cases, etc., these theories have themselves failed to satisfy the *methodological* standard of interpretation/idealization, the requirement, roughly, that the conception offered be specified with the precision appropriate for a physical interpretation of an abstract formal calculus, and be fully interpreted in terms of concepts, objects or phenomena understood independently of probabilistic concepts. This *may* be grounds for scepticism about objective probability. On the other hand, perhaps we should take seriously the idea that propensity theories are, in part, proposals of a *new metaphysical or physical hypothesis* and that, therefore, we should not expect propensities to be explicable, in the way the condition of interpretation/idealization demands, in terms of *old*, or independently understood, concepts, objects or phenomena. Perhaps, in view of the idea that propensities are supposed to be entities of a "*new metaphysical category*", it is inappropriate to foist the interpretation/idealization requirement on theories of propensity, since the requirement insists on explication of the proposed conception in terms of *old* ideas.

Indeed, in view of the foregoing discussion, it seems to me that the only way in which propensity theories can secure something like interpretation/idealization is *through* their conceptual adequacy, where the objects of interpretation, then, are such things as: theoretically adequate explanations of single events, and of physical regularities; causal laws; events and objects themselves; etc. (as is implicitly suggested by Fetzer 1981, pp. 295–296). Thus, if these things can be *identified* (by which I do not mean fully understood) prior to an understanding of propensities, and if a propensity theory of probability can characterize a role that a certain concept (i.e., the propensity concept) plays in these things, it will thereby have established something like "bridge principles" connecting the theoretical concept of propensity with the independently identifiable things listed above, thereby also giving an implicit or partial definition of the theoretical concept. It seems to me that it can only be in terms of satisfaction of a weaker kind of condition of interpretation/idealization, formulated in the light of these ideas, that the propensity concept can be identified, where whether or not such a mode of identification would be entirely adequate is not entirely clear.

Notes

[1] On Jeffrey's (1965) theory, only a *family* of pairs of probability and desirability functions is determined by a coherent set of preference data, where neither function is uniquely determined. On other theories, the subjective probability function is determined uniquely, but the desirability function is not.

[2] Of course this does not imply that the set of values of a sequence-function is infinite. Also note that the idea of limiting frequency, defined below, applies in the case in which the reference class is finite: "Notice that a limit exists even when only a finite number of elements x_i belong to [the reference class B]; the value of the frequency for the last element is then regarded as the limit. This trivial case is included in the interpretation and does not create any difficulty in the fulfillment of ... the ... axioms" (Reichenbach 1949, p. 72).

[3] Limiting frequencies aren't in general countably additive. See Van Fraassen (1979) on this and on the idea of limiting frequencies being defined on Boolean algebras.

[4] For example, in Cartwright (1979) and Skyrms (1980); for discussion of these and other such theories, and further references, see Eells and Sober (1983).

[5] My notation here differs from Kyburg's. Also, here, as in the sequel, the terms 'A', 'B', 'x', etc., are just that: terms of the relevant first order language. Sometimes, however, when it is

clear what the relevant model M is, I shall use just 'A', 'B', 'x', etc., as names for what they denote *in M*; at other times, I shall write '$R(A)$', '$R(B)$', etc., for the class or sequence which $M = \langle U, R\rangle$ *assigns* to 'A', 'B', etc., and $M(x)$ for what M assigns to an individual term x.

[6] It seems that '$p(a, b)$' should not be read as 'the probability of the singular event a (*happening*) . . .', but as 'the probability of a certain event's having the relevant attribute . . .,.

[7] Some philosophers have said that there is an ambiguity in Popper's writings in connection with whether his propensity theory is supposed to be a "long run" interpretation or a "single case" interpretation. In any case, in this section, I shall be considering the long run construal; in the sixth section of this paper, I consider a single case propensity approach.

[8] I owe the idea of this machine to Harry Nieves, who invented it to make a somewhat different point.

[9] Fetzer actually states the requirement in terms of *predicates* (rather than of "factors") and of reference class *descriptions* (rather than of the classes themselves).

[10] Of course one might say (as Fetzer has urged in private communication) that we actually do observe direct manifestations of single case propensities in each single event: namely, the occurrence or nonoccurrence of the relevant attribute. But, of course, such single case displays are inappropriate for the purpose at hand - namely, securing empirical interpretation for the *statistical* concept - since such single occurrences or nonoccurrences of the relevant attribute are, separately, completely uninformative in relation to the value of the single case propensity in question.

References

Baird D, Otte RE (1982) How to commit the Gambler's fallacy and get away with it. In: Asquith PD, Nickles T (eds) PSA 1982, Philosophy of science association, east lansing, Michigan, pp 169–180

Benacerraf P (1973) Mathematical truth. J Philos 70:661–679

Cartwright (N) 1979 Causal laws and effective strategies. Nous 13:419–437

Eells E (1982) Rational decision and causality. Cambridge University Press, Cambridge, England and New York

Eells E and Sober E (1983) Probabilistic causality and the question of transitivity. Philos Sci 50: 35–57

Fetzer JH (1971) Dispositional probabilities. In: Buck RC, Cohen RS (eds) Boston studies in the philosophy of science VIII (PSA 1970). Reidel, Dordrecht, pp 473–482

Fetzer JH (1974) Statistical probabilities: single case propensities vs. long-run frequencies. In: Leinfellner W, Köhler E (eds) Developments in the methodology of social science. Reidel, Dordrecht, pp 387–397

Fetzer JH (1977) Reichenbach, reference classes, and single case "Probabilities". Synthese 34:185–217. Errata, Synthese 37 (1978), 113–14. Reprinted in Salmon (1979)

Fetzer JH (1981) Scientific knowledge. Reidel, Dordrecht

Fetzer JH and Nute DE (1979) Syntax, semantics, and ontology: a probabilistic causal calculus. Synthese 40:453–495

Fetzer JH and Nute DE (1980) A probabilistic causal calculus: conflicting conceptions. Synthese 44:241–246. Errata, Synthese 48 (1981), 493

Giere RN (1973) Objective single-case probabilities and the foundations of statistics. In: Suppes P, Henkin L, Joja A, Moisil GC (eds) Logic, methodology and philosophy of science IV. North-Holland, Amsterdam, London, pp 467–483

Giere RN (1976) A Laplacean formal semantics for single-case propensities. J Philos Logic 5: 320–353

Hacking I (1965) Logic of statistical inference. Cambridge University Press, Cambridge, England and New York

Hacking I (1980) Grounding probabilities from below. In: Asquith PD, Giere RN (eds) PSA 1980. Philosophy of Science Association, East Lansing, Michigan, pp 110–16
Jeffrey RC (1965) The logic of fecision, 2nd edn (1983). University of Chicago Press, Chicago and London
Kyburg HE Jr (1974) Propensities and probabilities. Br J Philos Sci 25:359–375
Kyburg HE Jr (1978) Propensities and probabilities. In: Toumela R (ed) Dispositions. Reidel, Dordrecht, pp 277–301. This is a slightly revised version of Kyburg (1974)
Popper KR (1957) The propensity interpretation of the calculus of probability, and the quantum theory. In: Körner S (ed) Observation and interpretation in the philosophy of physics. Dover Publications, New York, pp 65–70
Popper KR (1959) The propensity interpretation of probability. Br J Philos Sci 10:25–42
Reichenbach H (1949) The theory of probability. University of California Press, Berkeley
Russell B (1948) Human knowledge: its scope and limits. Simon and Schuster, New York
Salmon WC (1967) The foundations of scientific inference. University of Pittsburgh Press, Pittsburgh
Salmon WC (1971) Statistical explanation and statistical relevance. University of Pittsburgh Press, Pittsburgh
Salmon WC (1979) Hans Reichenbach: logical empiricist. Reidel, Dordrecht
Settle T (1975) Presuppositions of propensity theories of probability. In: Maxwell G, Anderson RM Jr (eds) Minnesota studies in the philosophy of science VI: induction, probability and confirmation. University of Minnesota Press, Minneapolis, pp 388–415
Sklar L (1970) Is probability a dispositional property? J Philos 67:355–366
Skyrms B (1980) Causal necessity. Yale University Press, New Haven
Suppes P (1973) New foundations of objective probability: axioms for propensities. In: Suppe P, Henkin L, Joja A, Moisil GC (eds) Logic, methodology and philosophy of science IV. North-Holland, Amsterdam, pp 515–529
Suppes P (1974) Popper's analysis of probability in quantum mechanics. In: Schilpp PA (ed) The philosophy of Karl Popper. Open Court, La Salle, IL, pp 760–774
Van Fraassen BC (1979) Relative frequencies. In: Salmon (1979)
Von Mises R (1957) Probability, statistics and truth, 2nd English edn. Macmillan, New York
Von Mises R (1964) In: Geiringer H (ed) Mathematical theory of probability and statistics. Academic Press, New York

Part I
Alternative Conceptions of Probability

Probabilistic Causality and Causal Generalizations*

Daniel M. Hausman

Theorists of probabilistic causation have failed to distinguish between different tasks. One problem is to understand generalizations such as, "Smoking causes lung cancer," "Seat belts save lives," or "Just a spoon full of sugar helps the medicine go down." Some causal generalizations, like the examples I have just given, are immediately practical. Other causal generalizations, such as those that are central in economics may be more theoretical. Whether immediately practical or not, causal generalizations are problematic, because the cause they purport to identify are not invariably accompanied by their effects. They are in this way irregular.

As philosophers such as John Stuart Mill (1843) and, more recently, John Mackie (1980) have shown, such irregularity does not rule out the possibility that the underlying causal relations are deterministic. If a cause is a conjunct in a minimal sufficient condition for its effect, then the effect may fail to accompany the cause whenever any of the other conjuncts are absent. But why believe that there are minimal sufficient conditions for lung cancer involving smoking or for demand increases involving price drops? Why not formulate a theory of the probabilistic causality that is expressed in causal generalizations?

The fundamentally indeterministic relations identified by contemporary physics also seem to call for a theory of probabilistic causality. For example, the collision of a neutron with a uranium 235 nucleus raises the probability that the nucleus will decay, but it does not raise the probability to one. Contemporary physics tells us that such decay probabilities cannot be explained by underlying deterministic relations. Though some philosophers would deny that these indeterministic relations

D.M. Hausman (✉)
Herbert A. Simon Professor, Department of Philosophy, University of Wisconsin-Madison, 5197 Helen C. White Hall, 600 N. Park Street, Madison, WI 53706–1474
e-mail: dhausman@wisc.edu

*I owe a special debt to Ellery Eells and Elliott Sober, with whom I spent many hours discussing the issues raised in this essay, to Joonsung Kim, who devoted a chapter of his dissertation to criticism of an earlier version, and to Christopher Hitchcock for letting me appropriate so many of his ideas as well as for his criticisms of an earlier version of this essay. I am also grateful to Helen Beebee, Nancy Cartwright, Malcolm Forster, Huw Price, Peter Menzies, Charles Twardy, Jim Woodward, and John Worrall, who offered helpful comments on earlier drafts, to students in two seminars at the University of Wisconsin and to audiences at the University of Maryland, the London School of Economics, Duke University, Monash University, and Sydney University.

E. Eells and J.H. Fetzer (eds.), *The Place of Probability in Science*, Boston Studies in the Philosophy of Science 284, DOI 10.1007/978-90-481-3615-5_2,

are causal (Papineau 1989; Woodward 1989; Hausman 1998, ch. 9), most who have addressed the question maintain that contemporary physics reveals that there are indeterministic causal relations.

Philosophers have hoped that a single theory of probabilistic causality would account for both causal generalizations and the indeterministic relations identified by contemporary physics, though they have rarely attempted to extend their account to the causal generalizations of the special sciences such as economics. I shall argue that the issues raised by causal generalizations are largely independent of metaphysical questions concerning probabilistic causality. This argument does *not* suppose that the causal relations that underlie a claim such as "Smoking causes lung cancer" are deterministic. Whether the underlying relations are deterministic or not does not bear on the question of whether smoking causes lung cancer. Metaphysical theories of probabilistic causation should not be expected to provide truth conditions for causal generalizations or to guide us concerning how to make use of them.

Four Distinctions

Although this paper is mainly concerned to trace problems in theories of probabilistic causation to the mistaken assimilation of the issues raised by causal generalizations to those concerning indeterministic causal relations, there are other dimensions along which the objects of theories of probabilistic causality differ. Some of these are closely aligned with the difference between causal generalizations and claims about indeterministic causality, while others cut across this distinction. In particular, one should draw at least the following distinctions.

1. *Relevance versus role*. By the "role" or "bearing" of a causal factor, I mean whether a causal factor is positive or negative, or in some way "mixed" for some outcome. A poison and an antidote are both causally *relevant* to death, but their *roles* are opposite. A factor or variable X is causally *relevant* to another factor Y if it has *any* bearing on Y, positive, negative or mixed. Causal generalizations both in science and especially in daily life are typically concerned with role, rather than merely relevance. (They wouldn't be practical otherwise!) One wants to know whether smoking increases the risk of lung cancer, not merely whether it is somehow relevant. One wants to know whether lowering the price of a commodity will increase the demand for it, not just whether price changes are somehow relevant.
2. *Variables versus values of variables*. Consider a continuous variable Y that measures an agent A's income and a second continuous variable Q that measures the quantity of chocolate A demands. Y is causally relevant to Q, and over some ranges of values, one can say that Y has a positive impact on Q in the sense that as Y increases, so does Q. So one can speak meaningfully of both the causal role and causal relevance of one variable for another (given values of the other variables and within some range of values). Causal generalizations within a science such as economics usually concern or derive from claims about the role and relevance of variables. A claim such as the law of demand, for ex-

ample, follows from a more general functional relation postulated between the quantity demanded of a commodity, its price, income, and other variables. At the same time, one can also speak – as is more common in practical causal generalizations – of the causal role of the *value* of a variable X on some effect as a contrast between the effect of the given value and the effect of some other value of X that is pertinent in the context (see Hitchcock 1993, 1995, 1996). Smoking one pack of cigarettes a day increases the risk of lung cancer compared to not smoking any cigarettes, but it diminishes the risk compared to smoking two packs a day. The first contrast is of course usually the pertinent one. Practical causal generalizations typically concern relations among values of variables.

3. *Homogeneous versus heterogeneous circumstances*. Causal factors, as ordinarily conceived, have different consequences in different circumstances. Exposure to small pox will not cause the disease in those who have been inoculated. Suppose one had a list of all of the variables that are relevant to whether someone contracts lung cancer, other than S (the number of cigarettes smoked) and consequences of S. A generalization concerns causally homogeneous circumstances, if and only if these other variables have unchanging values. Practical causal generalizations (unlike claims concerning indeterministic causal relations within physics) typically concern causal relations when the values of other causally relevant variables are *not* unchanging. They concern causal relations in heterogeneous circumstances – that is, across some range of causally homogeneous circumstances.
4. *Types versus tokens*. Smoking can be a "type-level" cause of lung cancer – that is, it can tend to cause lung cancer – even when it does not actually do so. This difference between tendency and upshot cuts across the other distinctions.[1] In this paper I am concerned only with type relations – that is, with generalizations concerning causal tendencies. All of the other distinctions, including the distinction between causal relations in causally homogeneous circumstances and causal relations in heterogeneous circumstances are distinctions *among* causal generalizations, not distinctions between claims about actual causation and claims about causal tendencies.

If one simplifies and supposes that questions about causal relevance always arise with respect to variables rather than values of variables, then one can draw Table 1.

Table 1 Distinctions among causal generalizations

	Causal relevance (of variables)	*Causal role* of variables	of values of variables
Heterogeneous contexts	Demand is a function of prices, incomes, and other things	The cancer risk increases with the amount one smokes	Seat belts save lives
Homogeneous contexts	In circumstances K, neutron bombardment is causally relevant to decay	*Ceteris paribus* the demand for X is a decreasing function of the price of X	In circumstances K, $\Pr(\mathrm{Y} = \mathrm{y})/(\mathrm{X} = \mathrm{x}^*) > \Pr(\mathrm{Y} = \mathrm{y})/\mathrm{X} = \mathrm{x}')$

Clarifying these distinctions cuts through problems that have plagued theories of probabilistic causality. Theories that address the metaphysical question of what it is for a variable to be probabilistically causally relevant to some outcome belong in the bottom left-hand cell, while a metaphysical theory of causal role (if there is such a thing) belongs in bottom middle or the bottom right-hand cell. Practical causal generalizations *presuppose* that there are causal relations of some sort – whether deterministic or indeterministic – when the circumstances are causally homogeneous. Their job is to provide guidance when one does not know what the other causally relevant variables are and what their values may be. They are generalizations across homogeneous contexts. The puzzle they present is to say what sort of generalizations they are, not what causation is.

Before addressing these puzzles, let us use these distinctions to clarify the form of the causal generalizations. For definiteness, let us focus exclusively on the top right-hand cell – that is, on causal generalizations about the bearing of values of variables. Since the causal role of variables depends on background circumstances, causal generalizations should be relativized to some population P. Furthermore, claims about causal role always contrast the effect of one value of the purported causal variable to the effect of another value. This is trivial in the case of dichotomous variables, but in the case of non-dichotomous variables, the contrast should be made explicit. So I shall take the canonical form of a causal generalization to be:

$$\text{In population } P,\ X = x^* \text{ as compared to } X = x' \text{ causes } E.$$

In the case of dichotomous variables, this can be abbreviated as "In population P, C causes E," with the contrast between the effect of C and $\sim C$ understood. For practical purposes, one should require that the increase in the probability of E due to $X = x^*$ (or C) be substantial, but I shall leave this requirement implicit.

The Irregularity of Causal Generalizations

According to warning labels on cigarettes, the Surgeon General has determined that smoking causes lung cancer. What does this mean? Since not everyone who smokes gets lung cancer, smoking by itself is not a deterministic cause of lung cancer. But one need not surrender the view that causation is a deterministic relation. Similarly the fact that causal relations among economic variables are not invariable has not led economists to abandon a deterministic view of causation. Following Mackie (1980, ch. 3) one might say that deterministic causes are INUS conditions for their effects. Suppose that smoking were an INUS condition for lung cancer, and that price changes were INUS conditions for changes in quantity demanded (Hoover 2001, ch. 2). Although not by itself sufficient, smoking would be a conjunct in one or more minimal sufficient conditions for lung cancer. Since there are presumably also minimal sufficient conditions for lung cancer that do not include smoking, smoking would not be necessary for lung cancer either. If some of the relevant causal relations are not deterministic, then causes cannot be INUS conditions for their effects, but they can be conjuncts in minimal sufficient conditions that fix some objective chance of the effect occurring.

Let G be the other conjuncts in a minimal sufficient condition for lung cancer that includes smoking, and call the disjunction of the other minimal sufficient conditions for lung cancer "H." On a deterministic view of causation, smoking makes a difference to whether someone gets lung cancer only given the presence of G and the absence of H. Smoking is "necessary in the circumstances" – that is, necessary when none of the other minimal sufficient conditions for C are present, and sufficient when all of the other conjuncts in one or more of the minimal sufficient conditions including smoking are present. In the population as a whole, smoking has no single causal role. It causes lung cancer only in those individuals in whom just the right background conditions obtain. This account takes causation to be a three-place relation between a cause C, its effect E, and background conditions K in which C is necessary and sufficient for E. In a science such as economics, a good deal is known about the background conditions in which, for example, an increase in the money supply is necessary and sufficient for an increase in the rate of inflation, but the conditions cannot be completely specified.

Analyses such as these – whether deterministic or indeterministic – reveal a problem, which Wayne Davis calls "the background conditions problem" (1988, p. 133). The problem is that the third place in the causal relation means that causes are only causes when the conditions are "right." One can avoid introducing a third place in the causal relation by quantifying existentially: C causes E only if there exist background circumstances in which C is necessary and sufficient for E (and further conditions to insure the asymmetry of causation are met). But without saying more and without knowing whether in the actual circumstances C is necessary or sufficient for E, one is left with a very weak notion of causation. To say that smoking causes lung cancer is surely to say more than that there are some circumstances in which smoking is a conjunct in a minimal sufficient condition for lung cancer. Though people can rarely specify precisely what the other conjuncts in the minimal sufficient conditions including the particular cause are or what other minimal sufficient conditions there are for the given effect, they nevertheless usually know important facts about what background conditions must obtain for C to cause E. For example, without knowing exactly the conditions in which striking matches is necessary and sufficient for them to light, most people know that matches need to be dry.

Since people do not have detailed knowledge of the background conditions, why suppose that there *are* any minimal sufficient conditions for lung cancer or increases in the rate of inflation? Why should one believe that causation really is deterministic? Faith that causation is deterministic seems not only unjustified, but pointless as well, because it leaves one unable to say anything except that for some people smoking causes lung cancer, while for others it is irrelevant, and for still others it may prevent lung cancer. One might argue that only a dogmatic attachment to a deterministic theory of causation lends credibility to such a vague and unhelpful account. Why not focus directly on relations that people can know something about, such as the non-deterministic manifest relation between smoking and lung cancer in the actual inhomogeneous circumstances in which people live, smoke, and die?

Patrick Suppes pursues very much this line of thought.

> [A] mother says, "The child is frightened because of the thunder", or at another time, "The child is afraid of thunder". She does not mean that on each and every occasion that the child hears thunder, a state of fright ensures, but rather that there is a fairly high probability of its happening....
>
> It is easy to manufacture a large number of additional examples of ordinary causal language, which express causal relationships that are evidently probabilistic in character. One of the main reasons for this probabilistic character is ... we do not explicitly state the boundary conditions or the limitations on the interaction between the events in question and other events that are not mentioned.... A complete causal analysis is far too complex and subtle, and not to the point for which ordinary talk is designed. (1970, pp. 7–8)

Although Suppes here emphasizes the supposed conformity of probabilistic causality to ordinary language,[2] rather than the avoidance of a metaphysical commitment to determinism, he is following the line of thought sketched in the previous paragraph. He suggests that one can avoid invoking unknown minimal sufficient conditions by developing a theory of probabilistic causality.

The Surgeon General obviously means to say more than that smoking has some probabilistic relevance to lung cancer, be it positive or negative, and monetarists are claiming more than that the money supply has some relevance to inflation. We are warned that smoking significantly *increases* the probability of lung cancer or that increasing the money supply will lead to a non-trivial increase in the rate of inflation. So Suppes attempts to formulate a theory of what I have called "causal role."[3] One can do so without referring to some set of unknown conjuncts in minimal sufficient conditions, by maintaining that smoking causes lung cancer only if the probability of lung cancer conditional on smoking is larger than the probability conditional on not smoking *regardless of the circumstances*. Instead of a three-place relation between C, E and background conditions B, perhaps one can get rid of the intangible third place in the relation and analyze causation in terms of a two-place relation of statistical relevance. With the additional stipulation that smoking precedes lung cancer, this is basically Suppes' definition of a *prima facie* cause (1970, p. 12). Theories of probabilistic causation attempt in this way to dodge the background conditions problem.

Contextual Unanimity

But one cannot simply forget about the circumstances and other causal factors. It may be that smoking is not a positive cause of lung cancer even though $\Pr(C/S) > \Pr(C/{\sim}S)$. As R.A. Fisher postulated, some genetic common cause of smoking and lung cancer could explain the correlation. Neither is it necessary that $\Pr(C/S) > \Pr(C/{\sim}S)$. Some gene that makes people likely to smoke might impede lung cancer so that smoking and lung cancer are not positively correlated, even though smoking causes lung cancer. The probabilistic relations between smoking and lung cancer may be misleading.

The solution to these difficulties adopted by most theorists of probabilistic causation has been to require that smoking increase the probability of lung cancer within *all* cells of a partition formed by taking into account all the other causes of lung cancer that are not themselves effects of smoking.[4] The cells in such a partition are causally homogeneous background contexts, and the proposal can be restated as the requirement that causes increase the probability of their effects in all causally homogeneous background contexts. John Dupré has dubbed this the requirement of "contextual unanimity" (1984). Since theorists of probabilistic causality quantify over causally homogeneous background contexts, they avoid reintroducing a third place in the causal relation. Rather than relativizing causal claims to specific causally homogeneous contexts, theorists of probabilistic causation maintain that C is a positive cause of E (in some population P) if and only if C increases the probability of E in *every* causally homogeneous background circumstance in P (and some other condition is met that guarantees causal asymmetry, such as temporal priority of the cause).

Eells and Cartwright take analyses such as these to constitute a metaphysical theory of what causation is. In my view, in contrast, such theories are an amalgam of metaphysics and methodology. They combine a metaphysical theory of probabilistic causation within individual causally homogeneous background circumstances and a view of how causal generalizations generalize across such circumstances. Although the details vary, the implicit metaphysical view is that in some causally homogeneous circumstance, C is causally relevant to E if and only if it is probabilistically relevant to E, and C precedes E. The causal generalization, "C causes E" is then taken to maintain that C causes E within *every* causally homogeneous background circumstance.

I thus suggest that *contextual relativity has nothing to do with the metaphysics of causation*. The metaphysics in theories such as those of Suppes, Cartwright, Humphreys (1989), and Eells consists of the claim that causation consists of statistical relevance and temporal priority of the cause, given some particular value for each of the other causally relevant variables (that is, within individual causally homogeneous circumstances). Contextual unanimity figures instead in the attempt to explain how causal generalizations can be true and useful. In particular, contextual unanimity is the easiest way to avoid relativizing causation to particular contexts: if there are any circumstances (or "subpopulations" in Eells' terminology – 1991, ch. 1), in which smoking does not increase the probability of lung cancer, then smoking is not a cause of lung cancer in the population as a whole. By insisting on contextual unanimity, one is thus able to say more than merely that smoking increases the probability of lung cancer in some circumstances, though it may lower it in others, and be irrelevant in still others.

Two additional considerations motivate the requirement of contextual unanimity and the unwillingness to relativize causal claims to particular background circumstances. First, contextual relativity makes it easier for theorists to convince themselves (erroneously) that their accounts of causal generalizations are part of a metaphysical theory of probabilistic causal relations. If instead one concluded smoking could be said to be a cause of lung cancer only with respect to some contexts and not with respect to others, then the truth of unrelativized causal generalizations, such as the Surgeon General's, would depend on there being an implicit specification of background contexts. But such a specification of favored background contexts has no place in a theory in a metaphysical account of what causation is. By insisting on contextual unanimity, the difficulty vanishes: there is no need to justify zeroing in on some contexts and ignoring others.

Second, scientists do not know what the causally homogeneous background contexts are against which smoking may cause lung cancer, and even if these contexts were known, individuals do not know which cell of the partition they occupy. The Surgeon General needs to offer advice that is applicable to people who are not all in the same causal circumstances. If smoking had the *same* bearing on lung cancer in every cell of the relevant partition of other causal factors, then the Surgeon General could warn people about the risks of smoking without knowing anything about their particular circumstances.

The requirement of contextual unanimity is thus an attempt to *evade* the irregularity of causal generalizations. It only comes into play if one attempts to generalize across different causally homogenous background circumstances. It has no relevance to claims concerning the causal relevance of X to Y or the causal bearing of a value of X on Y within a single causally homogeneous background circumstance, which is where all the metaphysical action, so to speak, lies.

But the irregularity of causal generalizations cannot be evaded. Since, as the INUS analysis reveals, there is no reason to expect that a deterministic cause C of E will have the same bearing on E in every causally homogeneous circumstance, why should one stipulate that probabilistic causes must satisfy contextual unanimity? If probabilistic causes are INUS conditions for some objective chance of their effects occurring, one would expect them to be as sensitive to the background circumstances as deterministic causes; and there seems to be no general reason to suppose that the objective chance of their effect occurring will be increased by the presence of the cause in every homogenous context.

There are more specific grounds to doubt contextual unanimity. Consider a question posed by John Dupré: Should one conclude that smoking does not cause lung cancer if it were discovered that some people have a rare physiological condition that causes them to contract lung cancer more often if they do *not* smoke (1984, p. 172)? Indeed, no hypothetical case is necessary: smoking does not *in fact* increase the probability of lung cancer in every causally homogenous background situation. For example, in some people smoking causes fatal heart attacks rapidly enough that it tends to prevent lung cancer. Although smoking would be an undesirable way for these people to prevent lung cancer, it would do the job.

If one requires contextual unanimity, one thus has to deny that smoking is a positive cause of lung cancer, or one has to restrict the population to which the surgeon general's causal generalization is supposed to apply (Glennan 2002, p. 124). In just the same way, it turns out that seat belts don't save lives. Brushing one's teeth doesn't prevent tooth decay. Caffeine doesn't wake people up. Increases in the money supply don't spur inflation. Aspirins don't alleviate headaches. And, I suspect that a spoonful of sugar does not help this bitter medicine go down. In short just about every causal generalization turns out to be false, unless one radically restricts its scope.

Ellery Eells is willing to bite the bullet and to conclude that smoking does not cause lung cancer. He maintains that it is instead "causally mixed" for lung cancer (1991, p. 100). In some cells of the partition it increases the probability of lung cancer and in some cells it does not. Imposing a requirement of contextual unanimity implies that causes are in fact typically causally mixed for their purported effects. Given the contextual unanimity requirement, the only truthful causal generalization the Surgeon General can make about the consequences of smoking for Americans in general is that sometimes it causes smoking and sometimes it doesn't. But the Surgeon General is neither uttering this useless truth nor is he falsely maintaining that smoking increases the probability of lung cancer in every causally homogeneous context. In failing to capture what claims such as the Surgeon General's mean, theories such as Eells' are unable to distinguish useful and apparently true generalizations such as "Smoking causes lung cancer" or "Seat belts save lives" from useless and apparently false generalizations such as "Vitamin C cures cancer." Contextual unanimity is self-defeating in the analysis of causal generalizations and irrelevant to the metaphysics of indeterministic causation (see also Woodward 1989, p. 374).

One can try to save the contextual unanimity requirement by hedging causal generalizations or restricting their scope. Presumably there is some condition H in which it is true that seat belts invariably increase the probability of surviving crashes. But without knowing what H is or having any idea whether H will obtain in the event of an accident that might befall me, the true restricted generalization gives me no guidance concerning whether to wear my seat belt. To save the contextual unanimity analysis of probabilistic causal generalizations in this way is to make these generalizations useless.

So probabilistic theorists who insist on contextual unanimity are no more successful than the deterministic theorist in analyzing claims such as "smoking causes lung cancer." C can be a cause of E even though its bearing on E differs in different causally homogeneous circumstances. To interpret the causal generalization "C causes E in population P" as maintaining that C increases the probability of E in every homogeneous circumstance in this population implies that causal generalizations are almost all false or else have such a narrow or unclear scope as to be useless. Some other way to generalize across contexts is needed.

The right way to interpret causal generalizations is, I think, basically John Dupré's. Dupré's idea (which is developed more precisely by Eells (1987, pp. 108–110) and especially by Hitchcock (1998, pp. 282–290) is that one should hold fixed the frequencies of all the other factors relevant to lung cancer (apart from smoking and its effects) at their frequency in the actual population and see whether, against this background, the conditional probability of lung cancer, given smoking, is larger than the conditional probability of lung cancer, given non-smoking. Of course, if one knew what the causally homogeneous circumstances were, the role of the causal factor in each of those circumstances, and which circumstances individuals were in, then there wouldn't be any need to do any averaging. So one should resist interpreting Dupré as calling for people to *construct* these averages from more detailed knowledge. One can instead learn the average effect from comparing outcomes in treatment and control groups in randomized experiments or by inferences from observed correlations. On Dupré's construal, the Surgeon General's claim aims to provide just the information needed to decide whether to smoke by people who do not know how their propensity to develop lung cancer differs from the population average.

Although the truth of causal generalizations may depend on the relative frequencies of different causally homogeneous contexts, Dupré is not reducing causation to mere correlation. On Dupré's and Hitchcock's view – at least as I understand it – the generalization, "In population P, C is a significant cause of E" is true if and only if *in an ideal randomized experiment* the frequency of E would be appreciably larger among subjects taken from P who are exposed to C than among subjects taken from P who are exposed to $\sim C$. Although in general one would expect a correlation between C and E in the population, a correlation is neither necessary nor sufficient for an "average effect." A correlation is not sufficient, because it might reflect the fact that C and E are effects of a common cause. In such a case the existence of the correlation would not underwrite action to bring about C, and causal generalizations are, of course, supposed to guide action. One cannot prevent a storm by putting a barometer in a pressure chamber and thereby preventing its reading from falling. The existence of a correlation is not necessary, either. It could be that C causes an increase in the chance of E at the same time as some common cause counteracts this correlation. For example, if those who live in rural areas where other causes of lung cancer are absent are more likely to smoke, there might be no correlation between smoking and lung cancer, or even a negative correlation, even though both those living in rural areas and those living in urban areas are more likely to get lung cancer if they smoke.

Eells criticizes Dupré's proposal because it implies that whether smoking causes lung cancer depends on the actual frequencies of other factors. Change those frequencies, and smoking may cease to be a cause of lung cancer. In Eells' view, like Cartwright's, "smoking causes lung cancer" is supposed to be a causal law, which should not depend on the actual frequency of background conditions. So Dupré's account is "a sorry excuse for a causal concept" (Eells and Sober 1983, p. 54). Dupré agrees that his view makes laws depend on frequencies (1984, p. 173), but argues that this implication should be accepted. I disagree. Among other undesirable consequences, Dupré's view implies that people can change causal laws. Eells's critique of Dupré would be decisive, if the task were to formulate a metaphysical theory of indeterministic causation or to develop an account of probabilistic *laws*.

But both Eells and Dupré treat two tasks as one. Each sees that the other's theory is inadequate to the task with which each is mainly concerned. Dupré's theory is inadequate as a theory of indeterministic causation, which is what Eells is mainly concerned with. Eells' theory is inadequate as a theory of causal generalizations, which is what Dupré is mainly concerned with. The situation resembles that of two carpenters, one of whom mainly pounds nails, while the other more often screws things together. Both believe that a good carpenter needs only one tool. So the first uses only a modified hammer and the second only an odd screwdriver. The first points out how badly the screwdriver drives nails, while the second points out how badly the hammer turns screws.

What is at issue in theorizing about causal generalizations is causal irregularity. The operation of causal factors, whether deterministic or indeterministic, varies from context to context, and guidance is needed when the details concerning the contexts are not known. Theoretical work may focus on individual contexts or homogeneous contexts, because it need not necessarily provide such guidance. But if one hopes to offer advice to people who do not know which homogeneous context they are in, one has to generalize across contexts in which the effects of causal factors are not uniform. The point of the Surgeon General's claim is to provide information about the dangers of smoking to people who are in many different circumstances and who do not know which causally homogeneous context they are in. A generalization such as "smoking causes lung cancer" summarizes the qualitative "average effect," and it consequently depends not only on the cancer-causing propensities of smoking in causally homogeneous background contexts but also on the actual frequencies of the contexts.

Eells takes issue with this line of thought and argues that Dupré's account of irregular causation leads to mistaken advice. He writes,

> [T]he question of whether smoking is a population-level cause of lung cancer will turn on the population frequency of that physiological condition, and in an unacceptable way . . .. For example, a person contemplating becoming a smoker, and trying to assess the health risks, should not be so concerned with the population frequency of that condition, but whether or not *he* has the condition. That is, the person should be concerned with which *subpopulation* he is a member of, the subpopulation of individuals with the condition (a population in which smoking is causally negative for lung cancer) or the subpopulation of individuals without the condition (a population in which smoking is causally positive for lung cancer. (1991, pp. 103–104)

If, as Eells imagines, one knows the causal bearing of smoking on lung cancer in subpopulations in which contextual unanimity holds *and* one can find out which subpopulation one is in, then one should make use of the more specific information. So, for example, if smoking raised the probability of lung cancer in men but lowered it in women, then the Surgeon General's claim about the effect in the population as a whole, even if true, would be misleading. Rather than averaging across contexts in which smoking is positive, negative, or neutral for lung cancer, we should focus on its causal bearing in the context or subpopulation in which we find ourselves.

Dupré's and Hitchcock's formulations may misleadingly suggest that one begins with knowledge of the relevant causal factors and thus with a complete partition into causally homogeneous background contexts. One then determines the quantitative bearing of smoking on lung cancer in each of these contexts and the frequency of each context and thereby calculates the average effect of smoking on lung cancer. If this were an accurate description of the problem, Eells would be right to maintain that we should focus on the bearing of smoking on lung cancer in the contexts in which we find ourselves rather than averaging (though if we knew the particular context, there would be no need for generalizations across contexts and hence no need to impose a contextual unanimity condition either).

But this is not an accurate description of the problem. Nobody knows all the causal factors that are relevant to whether somebody contracts lung cancer. There is no way to calculate an average effect by summing over the effects in causally homogeneous circumstances weighted by their frequencies. Instead one can infer the average effect by means of experiment or critical examination of observed correlations. The point of the thesis that causal generalizations state average effects lies in justifying drawing causal conclusions from experiments and observations. These conclusions are important because agents who have no evidence about what subpopulations they belong to or concerning how the risks of lung cancer, for example, vary across different subpopulations can do no better than to rely on the average effect of smoking in the population as a whole.

Most theories of probabilistic causality fail to cope with the problems of causal generalizations, because these theories misconstrue the problems as calling for a metaphysical theory of probabilistic causality. They wind up either with metaphysical views that are hopeless as accounts of causal generalizations or with accounts of causal generalizations that are hopeless as metaphysical theories of causation. When considering claims such as "Seat belts save lives," knowing that there are subpopulations in which C is a cause of E – whether deterministic or indeterministic – is not to the point. What one wants to know is the causal significance of C for E when it is already suspected that C is "causally mixed" for E. Causal generalizations are supposed to help out here. Some do and some do not.

When Are Causal Generalizations True and Useful?

On the average-effect interpretation presented in the previous section, a causal generalization such as "In population P, $X = x^*$ as compared to $X = x'$ causes E'' is true if and only if (a) in P $\Pr(E/X = x^*) > \Pr(E/X = x')$ and (b) the probability difference in (a) is due to the causal influence of $X = x^*$ as compared to $X = x'$ in some causally homogeneous circumstance occupied by members of population P. (a) and (b) give truth conditions for causal generalizations concerning populations occupying causally heterogenous circumstances in terms of generalizations concerning causal relations obtaining within particular causally homogeneous background circumstances. Theorizing about these latter relations is a task for metaphysics. Since the task is to elucidate causal generalizations, rather than to clarify the nature of causal relations, one can help oneself to whatever theory of causation one prefers, provided that it preserves some link between causation and probabilities.

On this account (in contrast to Hitchcock 2001, pp. 219–220), causal generalizations can be true, yet useless or even seriously misleading. Suppose, for example, that eating French fries causes heart attacks among men in some circumstances and prevents heart attacks in women in some circumstances. At the level of the whole population, it turns out that eating French fries consequently increases the risk of heart attacks among men and within the population as a whole, but eating French fries lowers the risk of heart attacks among women. If the Surgeon General knew these facts and then announced only that eating French fries makes heart attacks more likely, he or she would be culpably misleading. But provided that the correlation really is a consequence of causal relations in causally homogeneous circumstances between eating French fries and heart attacks, this claim would be true.

This truth condition preserves and supports the intuition that part of the explanation for why some causal generalizations are useful is that they are true. But as just pointed out, a causal generalization can be true and misleading and even harmful to a great many people. If there were a dozen significant subpopulations in which the causal facts concerning the relationships between values of X and some effect E differed wildly, and everybody in the population P knew the facts about the subpopulations and knew which subpopulation he or she belonged to, then the average effect in P would be of little interest.[5]

There is, I believe, a great deal to be said about when practical causal generalizations are worth making, but little of philosophical interest. Clearly "C causes E in population P" is more worthwhile when $\Pr(E/C) - \Pr(E/{\sim}C)$ is large rather than small. It is more useful when E is more important. It is more useful when people are better able to bring about or to prevent C. True causal generalizations will in general be more useful than false generalizations, though falsehoods can, of course, sometimes have good consequences. For example, a mistaken causal claim that smoking causes acne could serve teenagers well by leading them to stop smoking and thereby to avoid heart attacks and lung cancer later in life. But the fact that falsehoods may do good is usually a bad reason for enunciating them. Finally, if the consequences of C for E differ appreciably over different subpopulations, then it can be harmful to generalize over the whole population. It is usually better to generalize concerning the narrowest populations for which the information is available.

It is also difficult to say much of philosophical interest about how individuals should change their behavior when they come to believe a causal generalization. One possibility, which has been developed carefully by Christopher Hitchcock (1998, 2001) is to idealize and suppose that the agent can estimate the subjective probability that he or she is located in each causally homogeneous background context and that causal generalizations provide the agent with knowledge of the difference that values of X make to the probability of E in each context. Knowing his or her own preferences, the agent can then choose the action that maximizes expected utility. More ordinary cases when the agent has little idea what the homogeneous contexts are or which he or she may be in can be modeled as cases in which the agent's expectation of the effect of $X = x^*$ on the chance of E will coincide with the average effect in Dupré's and Hitchcock's sense.[6]

I am skeptical about this approach because of the extreme idealizations it requires. What I prefer to say is simpler. Suppose that an individual agent A belongs to some population P for which it is true that C causes E, and that there is no narrower population to which A belongs for which there is any information concerning whether C causes E. Then A should regard actions that cause C as increasing the probability of E (in accordance with the generalization), unless A has some reason to believe that he or she belongs to some subpopulation of P in which C does not cause E. (For example, even though smoking causes lung cancer, those on death row in Texas probably do not have to worry about contracting lung cancer if they smoke.) Since, by assumption, agents are seeking guidance concerning what to *do*, specifically causal information is crucial. What matters to agents are the consequences of acting and bringing C about or preventing C, not whether $\Pr(E/C) > \Pr(E/{\sim}C)$. So what is useful to agents are specifically causal generalizations, not claims about mere correlations.

Consider, for example, the generalization, "Seat belts save lives." In the population of drivers in the United States, the probability of surviving accidents if one wears seat belts is significantly larger than the probability of surviving if one does not wear them (though, of course, people can argue about how "significant" the difference is). No doubt those who wear seat belts are on average more conservative drivers, and so some of the correlation between seat belt use and survival could be due to this common cause. But this common cause does not explain the correlation between seat belt use and survival among those who are in particular classes of accidents, and our knowledge of the mechanics of accidents supports the claim that seat belts really do save lives. "Seat belts save lives" is a true causal generalization.

This generalization is, moreover, worth formulating and in general worth acting on. This is so, even though there are certain classes of unusual accidents in which one is *more* likely to die if one is wearing a seat belt. That means that in some subpopulations the correlation is reversed, and this reversed correlation is equally a causal matter. If agents knew in advance which class of accidents they would be in, then the facts about the average effect of wearing seat belts in the whole population of drivers would be irrelevant. But before accidents occur, when the decision about whether to buckle one's seat belt must be made, there is no basis to assign individuals to the subpopulation of drivers who will be in those rare accidents in which seat belts diminish the odds of survival. A great many people consequently wear their seat belts (as they rationally should), because they believe that the causal generalization about the whole population grounds an expectation that they will be less likely to be injured or killed if they wear their seat belts.

Consider one more example. What should one say about the generalization, "Entering a hospital for treatment makes people more likely to die." Here there is a very significant correlation. The probability of death in the near future is much higher among hospital patients than among those who are not in the hospital. The problem with this generalization is that the increased mortality is not an average *effect* of going into the hospital. Although hospitals do kill people, the main explanation for the correlation is, of course, a common cause that sends people to the hospital for treatment and then kills them. So though there is a correlation, this causal generalization is false and does not provide a reason not to enter a hospital. There may be subpopulations, however, in which the correlation between death and hospital treatment is due to the dangers hospitals pose. For people with minor ailments whose local hospitals are exceptionally poor, the causal generalization, "Among people in this area with minor ailments, entering the hospital for treatment makes one more likely to die", could be true and useful.

Conclusions

To understand causal generalizations, one must understand how and why people generalize. Metaphysical theories of indeterministic causation need not trouble themselves with such questions. The metaphysical task is to clarify the causal relevance of variables within homogeneous contexts. Theorists of probabilistic causation tried to accomplish this task at the same time as they undertook to provide truth conditions for causal generalizations. They offered a probability increase and temporal priority view of causation within causally homogeneous background contexts, and they imposed a contextual unanimity condition to specify when causal generalizations are true. They then ran these two theories together into the view that causation is statistical relevance in all causally homogeneous circumstances (plus temporal priority of the cause). But the two theories should be pried apart. I have offered no assessment here of the view of indeterministic causation as statistical relevance plus temporal priority within a given causally homogeneous circumstance. Whatever one thinks of it, it is separate from the contextual unanimity account of causal generalizations, which I criticized.

The central point is that at least two theories are called for rather than one. In attempting to address at the same time all six of the cells in Table 1, near the beginning of this chapter, probabilistic theories of causation have wound up failing at their tasks. They offer no solution to the conundrums of practical causal generalizations, because they collapse in the typical case where the causal factors are mixed in the population as a whole, and one cannot specify in any non-trivial way subpopulations or circumstances in which contextual unanimity is satisfied. At the same time, they obfuscate and complicate indeterministic causation by focusing on problems that have little to do with the metaphysics of indeterministic causal relevance. There is no single relation of "probabilistic causality" manifested in quantum physics, everyday practical generalizations, and the causal claims of special sciences such as economics. The attempt to tackle both these problems in a single theory is a mistake and should be abandoned.

Notes

[1] Christopher Hitchcock argues that the type-token distinction confounds two distinctions: the distinction between claims about causal tendencies versus causal accomplishments ("actual causation") and the distinction between the scope of claims of both sort (2001, pp. 219–20).

[2] The claim to match ordinary usage is questionable. Notice that Suppes attributes to the mother the claim that causes make their effects highly probable, which conflicts with his own theory of causation as probability *increase*. In the same empiricist spirit, Wesley Salmon argues that one should focus on statistical rather than deterministic relations, because statistical relevance relations constitute the evidence for claims concerning irregular causal bearing (1984, pp. 184–185).

[3] One also needs at least a semi-quantitative theory to distinguish significant from unimportant causes, but I will ignore these problems in this essay. I believe contrasts similar to those I shall

discuss with respect to the theory of causal bearing play an essential role in providing a quantitative account of causal role.

[4] It is not easy to define the relevant partition precisely. For an early, influential, but flawed account, see Cartwright 1979, p. 26, and for criticisms of the details of her account, see Ray 1992, pp. 231–240 and Hausman 1998, p. 198. Such views abandon any attempt to offer a reductive analysis of causation in terms of probabilities. An alternative proposal defended by Brian Skyrms (1980) is to require only that causes increase the probability of their purported effects in some cells of the partition and that they never decrease the probability. This view is subject to the same criticisms as the requirement of contextual unanimity.

[5] An anti-drinking poster on a college campus proclaims, "Drinking causes AIDS". On the average-effect view, this claim is probably true. Many people are inclined to judge it to be false on the grounds that drinking does not bear the right *kind* of causal connection to a disease such as AIDS. Unlike sharing needles, a shot of whiskey does not carry the virus. Although "Drinking causes AIDS" misleadingly suggests a certain kind of causal connection, there is no need to build these domain-specific details into the truth conditions.

[6] Given the idealizations in this approach, this requires that the subjective probability agents assign to their occupying any particular causally homogeneous background context match its actual frequency and that the subjective conditional probability of the effect given the value of the causal variable in each context match the objective probability.

References

Cartwright N (1979) Causal laws and effective strategies. Noûs 13. Reprinted and cited from Nancy Cartwright. How the laws of physics lie. Oxford University Press, Oxford, pp 21–43

Davis W (1988) Probabilistic theories of causation. In: Fetzer J (ed) Probability and causality: essays in honor of Wesley C. Salmon. Reidel, Dordrecht, pp 133–160

Dupré J (1984) Probabilistic causality emancipated. Midwest Stud Philos 9:169–175

Eells E (1987) Probabilistic causality: reply to John Dupré. Philos Sci 54:105–114

Eells E (1991) Probabilistic causality. Cambridge University Press, Cambridge

Eells E, Elliott S (1983) Probabilistic causality and the question of transitivity. Philos Sci 50:35–57

Glennan S (2002) Contextual unanimity and the units of selection problem. Philos Sci 69:118–137

Hausman D (1998) Causal asymmetries. Cambridge University Press, Cambridge

Hitchcock C (1993) A generalized probabilistic theory of causal relevance. Synthese 97:335–364

Hitchcock C (1995) The mishap at Reichenbach falls: singular vs general causation. Philos Stud 73:257–291

Hitchcock C (1996) Farewell to binary causation. Can J Philos 26:267–282

Hitchcock C (1998) Causal knowledge: that great guide of human life. Commun Cogn 31:271–296

Hitchcock C (2001) Causal generalizations and good advice. Monist 84:218–241

Hoover K (2001) Causality in macroeconomics. Cambridge University Press, Cambridge

Humphreys P (1989) The chances of explanation. Princeton University Press, Princeton, NJ

Mackie J (1980) The cement of the universe. Oxford University Press, Oxford

Mill JS (1843) A system of logic. Longmans, Rpt. London 1949

Papineau D (1989) Pure, mixed and spurious probabilities and the significance for a reductionist theory of causation. In: Kitcher P, Salmon W (eds) Scientific explanation. University of Minnesota Press, Minneapolis. Minnesota Studies in the Philosophy of Science, 13, pp 307–348

Ray G (1992) Probabilistic causality reexamined. Erkenntnis 36:219–244

Salmon W (1984) Scientific explanation and the causal structure of the world. Princeton University Press, Princeton, NJ

Skyrms B (1980) Causal necessity: a pragmatic investigation of the necessity of laws. Yale University Press, New Haven, CT

Suppes P (1970) A probabilistic theory of causality. North Holland, Amsterdam
Woodward J (1989) The causal mechanical model of explanation. In: Kitcher P, Salmon WC (eds) Minnesota studies in the philosophy of science: vol. 8. Scientific Explanation, University of Minnesota Press, Minneapolis, pp 357–383

The Possibility of Infinitesimal Chances

Martin Barrett

A well-known puzzle about the relation between chance and possibility is illustrated by the fair spinner, a perfectly sharp, nearly frictionless pointer mounted on a circular disk. Spin the pointer and eventually it comes to rest at some random point along the circumference, which we identify with a real number in the half-open interval $[0, 1)$.[1] (Imagine a scale of numbers around the edge.)

In constructing a model for this device, physical symmetry guides our assumptions about chance and possibility:

(A) The chance[2] that the pointer initially at position $a \in [0, 1)$ will come to rest somewhere in a set $S \subset [0, 1)$ is independent of a. Let us write $X \in S; a$ for the event that the pointer initially at a comes to rest in S. (X is the random variable representing the final position of the pointer.) Then $\Pr(X \in S; a) = \Pr(X \in S; b)$.

(B) The possibility that the pointer initially at rest at position a will come to rest somewhere in a set S is independent of a. That is, $X \in S; a$ is possible if and only if $X \in S; b$ is possible.

(C) The chances and possibilities are unaffected if the scale is rotated (the scale affects no physical property of the spinner). This means that $\Pr(X \in S; a) = \Pr(X \in S - c; a - c)$. ($S - c$ is the set S translated to the left by an amount c.)

To these symmetry conditions, we may add:

(D) At least one point outcome is possible (i.e., there is some x for which it is possible that $X \in \{x\}$), because in any trial the pointer must stop somewhere.

(A) and (C) together entail that the probability distribution of X does not depend on the initial position (so we may write simply $\Pr(X \in S)$), and that the distribution is fully *translation invariant* (modulo 1). It is known that the only distribution compatible with this requirement is the uniform distribution on $[0, 1)$ also known as Lesbesgue measure on $[0, 1)$ or $\mathcal{L}$. When X is uniformly distributed, $\Pr(X \in [a, b]) = b - a$ and $\Pr(X = a) = \Pr(X \in \{a\}) = 0$.

(B) and (C) entail that all point outcomes are equipossible, either all possible or all impossible.[3] (B), (C), and (D) together entail that all point outcomes are possible.

M. Barrett
University of Wisconsin–Madison, Madison, Wisconsin, USA

E. Eells and J.H. Fetzer (eds.), *The Place of Probability in Science*, Boston Studies in the Philosophy of Science 284, DOI 10.1007/978-90-481-3615-5_3,

This last conclusion poses a dilemma, for it stretches the conception of possibility. Chance zero outcomes suffer from unlimited infrequency. Fix on any particular outcome, say $X = 0.37498190502$. Spin the pointer repeatedly; on the assumption of independent trials what is the chance that $X = 0.37498190502$ will occur at least once in 100 spins? In $100,000,000$? In a googolplex ($=10^{10^{100}}$) of lifetimes of the universe? Answer in each case: less than 10^{-100}, less in fact than any positive threshold, no matter how small. While chance zero outcomes may be possible, the possibility is utterly remote. This is why common sense and scientific practice both tend to treat such outcomes as de facto *im*possible, contrary to what (B), (C), and (D) entail.

The simplest response to the dilemma retains (D) but rejects the symmetry conditions (B) and (C). It recognizes as possible the actual outcome and only the actual outcome. This deflationary response collapses the possible onto the actual, in effect dispensing with possibility altogether. From this point of view, the metaphysical puzzle does not even arise.

Some philosophically inclined scientists or philosophers of science may look favorably on this response. Although probability in one form or another is insinuated into every branch of modern science, isn't possibility just a metaphysical flight of fancy? Perhaps, but possibility is a fundamental ingredient in the modern analysis of counterfactuals, which in its turn forms the foundation of at least one highly influential account (Lewis's) of cause and effect. So it is worthwhile to forego this easy response in order to see whether we can't better characterize the relation of the two concepts.

A second response retains (B) and (C) but rejects (D). Zaman (1987) takes the position that all point-outcomes should be regarded as impossible. He argues that we can consistently maintain the necessity of $X \in [0, 1)$ (equivalently, $\exists x(X = x)$) while denying the possibility of each instance $X = a$. We can do this if our logic is κ-inconsistent[4] for $\kappa = 2^{\aleph_0}$, the power of the continuum. His strategy bears some resemblance to supervaluationist analyses of vagueness, which explain how we can accept $A \vee B$ as true without having to accept either that A is true or that B is true. If I understand Zaman correctly, after a trial results in the outcome $X = a$, it is still the case that $X = a$ is impossible, but each disjunctive outcome $X \in (a - \epsilon, a + \epsilon)$ (for $\epsilon > 0$) is necessary.

I am sympathetic to Zaman's goal of providing a consistent analysis which would hold point-outcomes to be impossible, but the logical price is high. Why should we take our logic to be κ-inconsistent? Where is the modal semantics which would validate his modal propositions about the spinner?

The principal purpose of this article is to examine a third response to the dilemma. This response takes $\Pr(X = a)$ to be infinitesimal rather than zero, and so preserves the supervenience of the possibility of $X = a$ on the chance of $X = a$. This option seems to have been attractive to Lewis (Lewis 1973), who wanted to maintain the truth of the conditional "if $X = a$ has chance zero, then $X = a$ will not occur." As he puts it:

> Zero chance is *no* chance, and nothing with zero chance ever happens.... Infinitesimal chance is still *some* chance.

A terminology has arisen to describe distributions like this: a distribution of chances is *regular* if it "assigns a positive probability to each outcome it considers possible" (Elga 2004). If we can find an adequate regular distribution for the fair spinner, it would solve the metaphysical puzzle. And the use of such distributions would benefit other aspects of probabilistic methodology as well, for both classical and Bayesian statisticians find probability zero outcomes inconvenient. The former cannot, for example, form likelihood ratios with zero probabilities, and the latter cannot condition on learned propositions of probability zero. Both have devised workarounds, but the conceptual problems remain.

But can we find an adequate regular distribution for the fair spinner? I claim that we cannot.

The Standard Model

The description of the Fair Spinner given in the first paragraph is, of course, an idealization. Physically, there are no perfectly sharp pointers or nearly-frictionless mechanisms. If real spinning dials happen to exemplify genuine chance processes, the true probability distributions are doubtless lumpy or otherwise depart from the perfect uniformity of Lesbesgue measure. Nevertheless, even apart from instrumental considerations, there is more than one reason to take seriously the search for an adequate probabilistic model of the idealization.

One reason derives from the way in which idealized models may arise: as a limit of a sequence of other (perhaps more realistic) models. The fair spinner occurs as a limit in this way. Consider the discrete distributions D_n, in which the probability is concentrated at the points $0, 1/n, 2/n, \ldots, (n-1)/n$ for $n > 0$. (Each point has probability $1/n$.) D_n is the appropriate model for a roulette wheel with n cells. As n gets larger, it becomes increasingly difficult to distinguish D_n from $\mathcal{L}$ "with the naked eye". In fact, the sequence of discrete distributions $D_1, D_2, \ldots$ converges in the limit to the continuous distribution $\mathcal{L}$ in the technical sense called weak convergence.[5]

But the translation invariance which so perfectly captures the rotational symmetry of the fair spinner is not present in any of the D_n. If we considered only the D_n and not the limit, we would miss this critical characteristic. (If we rotate the scale by an amount $1/2n$, the probability distribution becomes concentrated on a completely disjoint set of points in [0,1).) The lack of translation invariance is an indication of the failure of D_n to adequately represent the fair spinner.

Another reason is that according to our best current theories, at least some physical quantities are not discrete or granular (like electrical charge) but are continuous and may take any value in an interval of real numbers. Where there are continuous magnitudes there may be continuous chance processes. The spatial position of a particle is such a magnitude, even according to quantum mechanics, and the outcome of a positional measurement of a particle is such a chance process. Even though there may be no actual fair spinners, there may be (for all we presently know) some actual continuous chance process whose distribution is isomorphic to our model for the fair spinner, and so what seems to be idealized in one context may turn out to be descriptive in another.

An adequate model should be constructed in accordance with the basic principles of probability and the symmetry principles (A) and (C). We already have one model which is adequate, except for regularity, and that is the standard model ($\mathcal{L}$).

The standard model is a *countably additive* probability measure on a σ-algebra of subsets of $[0, 1)$. In the setting of standard analysis, countable additivity is what makes probability theory go; without it we wouldn't have, for example, the Strong Law of Large Numbers, the Central Limit Theorem, or Kolmogorov's existence theorem for stochastic processes. But existing foundational accounts of probability are not strong enough to necessitate more than finite additivity, so requiring countable additivity of probability distributions is controversial. I shall simply require that an adequate replacement for the standard model be defined on (at least) an algebra of sets, and be (at least) finitely additive.

The standard model is invariant under translation (modulo 1) by any real number in $[0, 1)$, and this is entailed by the symmetry principles (A) and (C). An adequate replacement should meet the same standard, at least insofar as it should be translation invariant for any translation which does not alter the set of outcomes which it considers possible.

The standard model is *not* regular, since it assigns probability zero to every point outcome (technically, to every singleton $\{x\}$, for $x \in [0, 1)$). A regular model, using infinitesimals, cannot be an extension of the standard model, in the sense that it retains all the existing probability values assigned by the standard model. But an adequate model should approximate the standard model: the probability that the pointer stops in any interval $[a, b]$ should be at least infinitesimally close to $b - a$.

A model satisfying the approximation requirement will be empirically equivalent to the standard model, since no frequency data derived from observations can distinguish between chances that differ only by an infinitesimal. A constructive empiricist might argue that no more is needed to characterize adequacy. But the constructive empiricist has already dispensed with the need both for traditional modality (Van Fraassen 1980, p. 197) and chance (propensity), so for him or her the puzzle which guides this paper does not arise, as it does not for the deflationist mentioned above.

To take the metaphysical puzzle seriously is already to admit that there may be additional desiderata for adequacy beyond the approximation requirement.

Some Nonstandard Analysis

The use of infinitesimal real numbers is the domain of Nonstandard Analysis (NSA). Originated by Abraham Robinson in the 1960s, NSA was not the first rigorous treatment of infinitesimals (Schechter 1997, pp. 247–254), but it was the first to extend all of real analysis comprehensively, in such a way as to preserve all the first-order properties of the standard universe.

The literature is extensive, and there are different approaches. There is space here to give only an elementary exposition of a model–theoretic approach preferred by measure theorists (see Nelson 2001 for a popular axiomatic approach).

1. NSA operates with two structures, which we may call the *standard universe* and the *nonstandard universe*. Both are built up in levels. The standard universe, here denoted $\mathcal{U}$, contains all the ordinary real numbers at level one, all sets of real numbers at level two, all sets of entities from levels one and two at level three, and so forth through all finite levels. $\mathcal{U}$ is the union of all the levels. The level one entities are thought of as individuals; every other entity in $\mathcal{U}$ is a set. Using elementary set theory we can prove that this standard universe contains all relations of the reals, all real-valued functions of reals, all functions of functions, etc. In fact, it contains every mathematical entity normally thought essential to mathematical analysis and geometry.

   The nonstandard universe, denoted $^*\mathcal{U}$, contains at level one an *extension* of the set of reals which includes all the standard reals but also includes infinite reals, and numbers infinitesimally close to standard reals. (Infinitesimals are numbers infinitesimally close to 0.) The entities of this set are called nonstandard reals or *hyperreals*. Higher-level entities are then built up from lower-level entities just as in the standard universe, and $^*\mathcal{U}$ is the union of all finite levels.
2. Every standard entity $s \in \mathcal{U}$ (s may be a real number, a set, a set of sets, etc.) has a unique *counterpart* $^*s$ in the nonstandard universe, but not vice versa.[6] Within the nonstandard universe, the entities which are counterparts are also sometimes called standard entities and anything which is not a counterpart is called a nonstandard entity. The counterpart $^*\mathbb{R}$ of the set $\mathbb{R}$ of standard reals is *not* just the set of counterparts of standard reals. It contains all those counterparts plus additional nonstandard members (such as infinitesimals). In general, counterparts of infinite sets contain extra nonstandard elements.

   Note that despite what the notation suggests, $^*\mathcal{U}$ is not a counterpart, because $\mathcal{U}$ is not itself a standard entity; i.e., $\mathcal{U}$ is not a member of $\mathcal{U}$.
3. We can formulate sentences about the standard and nonstandard universes in a first-order formal language L which consists of the usual logical symbols $\neg, \wedge, \vee, \rightarrow, \leftrightarrow, \exists, \forall$, a name $\mathbf{s}$ for every *standard* entity s, and a single two-place relation symbol $\in$ denoting set membership. Formulas of L are displayed in boldface type. Two entity names which recur frequently below are $\mathbf{N}$ and $\mathbf{R}$, standing for the set of natural numbers ($\mathbb{N}$) and the set of real numbers ($\mathbb{R}$) respectively. The entity name $\mathbf{s}$ in a formula denotes the standard entity s when the formula is interpreted in the standard universe and the counterpart $^*s$ when it is interpreted in the nonstandard universe. So, for example, $\mathbf{R}$ denotes $\mathbb{R}$ in $\mathcal{U}$ and $^*\mathbb{R}$ in $^*\mathcal{U}$. Examples of simple sentences of L are $\boldsymbol{\pi} \in \mathbf{R}$ and $\forall \mathbf{x}(\mathbf{x} \in \mathbf{N} \rightarrow \mathbf{x} \in \mathbf{R})$. We usually write the second sentence in bounded[7] form, as $(\forall \mathbf{x} \in \mathbf{N})(\mathbf{x} \in \mathbf{R})$. Both of these sentences are true in the standard universe.
4. The crucial *Transfer Principle* says that every bounded sentence of L is true in the standard universe if and only if it is true in the nonstandard universe. The Transfer Principle is a theorem of NSA and one of the principal sources of its power. It entails, for example, that both of the simple sentences from the previous paragraph are true also in the nonstandard universe. By means of examples such

as these, the properties and relations of standard entities are "transferred over" to the nonstandard universe.

5. In the nonstandard universe, there is an important distinction between *internal* sets, which are covered by applications of the Transfer Principle, and *external* sets, which are not. Formally, a set S in the nonstandard universe is internal if and only if it is a counterpart or a member of some counterpart; it is external otherwise.

   For example, let $\mathbb{N} = \{x \in \mathbb{Z} : x > 0\}$ and $\mathbb{N}' = \{{}^*x : x \in \mathbb{N}\}$. $\mathbb{N} \in \mathcal{U}$ is the set of ordinary natural numbers. $\mathbb{N}' \in {}^*\mathcal{U}$ is the set of counterparts of ordinary natural numbers. The sentence[8] of L

$$\neg(\exists \mathbf{x} \in \mathbf{R})(\forall \mathbf{y} \in \mathbf{N})(\mathbf{x} > \mathbf{y}) \tag{1}$$

   asserts a truth about $\mathbb{N}$ in the standard universe, that no real number is bigger than every member of $\mathbb{N}$. In the nonstandard universe, however this is not true of $\mathbb{N}'$, because any infinite positive real is bigger than every member of $\mathbb{N}'$. Why doesn't this contradict the Transfer Principle? Well, Eq. 1 *is* true in the nonstandard universe, but there the symbol **N** denotes not $\mathbb{N}'$, but ${}^*\mathbb{N} = \{x : x \in {}^*\mathbb{Z} \wedge x > 0\}$, the counterpart of $\mathbb{N}$. ${}^*\mathbb{N}$ contains not only all the members of $\mathbb{N}'$, but also all the nonstandard positive integers, which includes the infinite ones.

   ${}^*\mathbb{N}$ is an internal set and has all the properties that can be expressed by sentences of L which are about $\mathbb{N}$. $\mathbb{N}'$ is an external set (this is proven in the appendix), and the sentences of L say nothing about it.

   NSA is almost exclusively about internal sets, mappings, measures, etc., because these are the entities we can "get ahold of." In other words, they are the entities of which we can deduce properties, using the Transfer Principle.[9] In fact, in the axiomatic approach propounded by Nelson (2001), external sets do not even exist; their formation by abstraction is termed "illegal set formation."

   In the model–theoretic approach summarized here, we may form external sets (mappings, measures, etc.), but we will find it difficult to say much that is useful about them, or to apply them in any significant way in a scientific theory.

   There is a useful necessary condition for internality which I apply below: a set is internal only if all its elements are internal. This entails that a nonstandard probability distribution is internal only if the space of outcomes, the algebra, and the probability mapping are all internals.

6. In the world of NSA, there arises an important new notion of finiteness, called *hyperfiniteness*. In ordinary (standard) mathematics, we often formulate the notion of finiteness by saying that X is finite if and only if it can be indexed by an initial segment $\{1, 2, \ldots, n\}$ of the natural numbers $\mathbb{N}$, so that it has n elements which can be displayed $x_1, x_2, \ldots, x_n$. The "indexing" is technically a 1–1, onto function from $\{1, 2, \ldots, n\}$ to X.

   This concept transfers perfectly over to the nonstandard universe, only here an initial segment of the natural numbers is a set $\{1, 2, 3, \ldots, M\}$, for $M \in {}^*\mathbb{N}$. Now M may be an infinite natural number. We say that $X \in {}^*\mathcal{U}$ is hyperfinite with M elements if and only if there is an internal indexing function from $\{1, 2, 3, \ldots, M\}$ to X (this forces X to be an internal set, because the domain and range of an internal function are internal). All ordinarily finite sets fall under this definition as well.

In applications of NSA, models for the nonstandard universe are always constructed to be what Robinson called *enlargements*. In an enlargement, for any standard set S there is a hyperfinite set which contains all the counterparts of elements of S. For example, (the counterparts of) all standard reals are contained in a single hyperfinite set. Thus hyperfinite sets can manage to be both small and large at the same time: small in the sense that any sentence of L true of finite sets is also true of hyperfinite sets by the Transfer Principle, and large in the sense that they may contain (externally) large subsets, like $\mathbb{R}$. Many of the important applications of NSA have concerned analysis with hyperfinite sets.

Nonstandard Models for the Fair Spinner

I now want to examine whether NSA will yield an adequate regular model for the fair spinner. Each candidate model I discuss is a probability distribution taking probability values in the hyperreal interval $^*[0, 1]$ (so infinitesimal probabilities are admissible.) We may suggestively think of a string (formalized as a one-dimensional manifold in space) whose points are in 1–1 correspondence with the hyperreal interval $^*[0, 1)$, and which is wrapped around the circumference of the spinner as a scale. There is a distinguished subset S of points marked on the string; these may be thought of as the points which the model "considers possible" (in Elga's characterization of a regular probability distribution), the points at which the pointer is "allowed" to stop. (S may be the whole interval $^*[0, 1)$.) To complete the picture, we specify the algebra of subsets of S and probabilities for each member of the algebra such that the probability measure is finitely additive. We explicitly assume that each single point $s \in S$ has some probability value (i.e., each singleton $\{s\}$ is in the algebra). This is reasonable, given that we regard the pointer as being perfectly sharp, i.e. "one point wide."

1. *The Hyperfinite Spinner* In this model, the string is marked at the M points in a large hyperfinite subset of $^*[0, 1)$, where M is an infinite natural number. In the simplest such model, the M points are evenly spaced at a distance of $1/M$, each with probability $1/M$. (There is even a set like this which contains (the counterparts of) all the rational numbers in [0, 1]).

   More generally, the points might be unevenly spaced (though infinitesimally close to each other), with differing probabilities. Where the points bunch up, the probabilities would be correspondingly reduced, so that the probability assigned to the interval $[a, b]$ could be infinitesimally close to $b - a$, as mandated by the approximation requirement.

   This is an attractive picture. To the "naked eye" (which we assume cannot distinguish between infinitesimally close points), the marked points would seem to cover the whole string. The distribution is regular and all the resources of finite, discrete probability theory are available to make calculations and predictions.[10]

But there are warts, for the hyperfinite spinner is full of unexplained arbitrary features. As a description of a chance process, it has all the same deficiencies as the ordinarily finite distributions D_n discussed above.

Why does S contain M points? Why not $M - 1$, or $M + 1$, or 2^M?

No matter how large M is, S is really just a negligible subset of $^*[0, 1)$, even though it may contain all the counterparts of standard reals in [0, 1]. One way to express its smallness is that by the Transfer Principle, when S contains M points, there are more than 2^M points of $^*[0, 1)$ which are not in S. Why do we consider that the pointer may not stop at any of these points?

And why does S consist of *these* M points rather than a different set of M points? If we denote the smallest nonzero point of S by a (there is such a smallest point by the Transfer Principle), and if we shift the whole hyperfinite set to the left by $a/2$, then there is at least one point in the shifted set that is not in S. Why not take S to be the shifted set instead?

The shifting by $a/2$ also shows that the model is not (generally) translation invariant.

Bartha and Johns (2004) reject an *epistemic* model because of a similar arbitrariness in the choice of an infinitesimal which characterizes the model. I think we should be inclined all the more to reject such arbitrary features in the present metaphysical context, the context of models for chance processes.

2. *The Sprinkle Spinner* Here the set S contains the counterparts of standard real numbers, and only those. The idea may be that since the standard model finds it sufficient to represent the circumference of the spinner by the ordinary (i.e. standard) real interval $[0, 1)$, this should suffice for a nonstandard model as well. We may imagine (counterparts of) the standard real numbers spread out along the string like sprinkles. Because the real numbers are dense (between any two reals there is another real), seen from a distance they appear to cover the whole string, just as in the hyperfinite spinner.

   This idea dates back to the early days of NSA. Bernstein and Wattenberg (1969) prove the existence of such a model which meets our approximation requirement. They prove the existence of a hyperfinite set F which has M points, contains the counterpart of every standard real in $[0, 1)$, and possesses as much translation invariance as can be gotten. Then they use the counting measure on this set to define a measure μ on all subsets of $[0, 1)$: $\mu(A) =$ (the number of points in $^*A \bigcap F)/M$. This entails that the measure of each singleton is $\mu(\{x\}) = 1/M$. They also prove that when A is Lebesgue-measurable, $\mu(A)$ is infinitesimally close to $\mathcal{L}(A)$.

   Skyrms (1980) seems to have been the first philosopher to have noticed this result and writes of it optimistically:

   > As Bernstein and Wattenberg have shown, there is a finitely additive, almost translation invariant, regular measure defined on all subsets of [0, 1]. Nonempty sets of Lebesgue measure zero, as well as [their translations], then receive infinitesimal measure. And we can say the probability of our pointer hitting a rational in $[0, 1/2]$, given that it hits a rational, is the ratio of the two appropriate infinitesimals and equals $1/2$.

As in the case of the hyperfinite spinner, $S = \{{}^*x : x \in \mathbb{R}\}$ is a negligably small subset of ${}^*[0, 1)$, but here at least one might defend this choice of S on the grounds that it is a distinguished (i.e., non-arbitrary) subset of ${}^*[0, 1)$. However, the model is *not* translation invariant, as Skyrms's "almost" lets out. It is invariant only under translations by standard rationals, and though this restriction may be by mathematical necessity, it is difficult to accept on ontological grounds. For translations by other standard reals, Bernstein's and Wattenberg's proof shows that the best we can get is that Pr(translated set) and Pr(original set) are infinitesimally close, not necessarily equal.

Because of this, the model does not give an adequate account of conditional probabilities over sets of infinitesimal probability either. The conditional probability relation mentioned by Skyrms in the quotation above does not hold generally. In the appendix is described a countable set X (which therefore has standard Lebesgue measure zero, like the rationals in $[0, 1/2]$) such that (a) X has an irrational translate Y which is disjoint from X and (b) $Pr(X|X \bigcup Y) \neq 1/2$.

The Bernstein and Wattenberg Theorem has served as a point of departure for a number of results in the philosophical literature asserting the existence of regular probability distributions, frequently in a rather abstract setting. See , for example, Mcgee (1994) or Elga (2004), in which the authors discuss regular nonstandard probability distributions which approximate standard ones up to an infinitesimal.

What these authors do not observe, however, is that like Bernstein's and Wattenberg's distribution, these models are all *external*. This is because they take an existing standard structure, like the σ-algebra of subsets of $[0, 1)$, or the σ-algebra of Lebesgue-measurable subsets of $[0, 1)$, or an infinite algebra of propositions, and map *that structure* (rather than its nonstandard counterpart) to ${}^*[0, 1)$. Such a mapping is external because its domain is external, containing only standard elements (all infinite sets containing only standard elements are external; see the appendix for a proof). And as I indicated above, if the probability mapping is external, the whole model is external.

I have no conclusive argument that a model for the fair spinner or other chance process cannot be an external nonstandard model. But such a model represents a big loss of instrumental virtue. In substantive applications of the theory of chance, there will typically be found constraints on a model arising from physical properties or symmetries in the physical situation. Because we cannot apply the Transfer Principle to an external nonstandard model, the task of showing that an external nonstandard model conforms to these constraints may become difficult or impossible. The lack of true translation invariance in Bernstein's and Wattenberg's model is an example of this.[11]

As an aside, the defectiveness of external models partly explains why nonstandard measure theory has developed in a direction quite different from that suggested by the early paper of Bernstein and Wattenberg. Subsequent research has focussed almost entirely on so-called Loeb measures, which are pairs consisting of an *internal* nonstandard measure (usually hyperfinite) and a *standard* measure (no infinitesimals) which it approximates. See Ross (1997) for an exposition of Loeb Measures and a discussion of the history. Loeb measures do not provide usable models for the present purpose because the standard part cannot be regular and the internal part is subject to the objections made to the hyperfinite spinner or else to the considerations of the next section.

3. *The Hyperreal Spinner* Finally, there is the class of *internal* models with $S = {}^*[0, 1)$ (all points on the string are marked). Since they allow that the pointer might stop at any hyperreal point, these models avoid unexplainable arbitrariness in the selection of points of S, and since they are internal, the Transfer Principle may be liberally applied to deduce their structure.

   The obvious first choice for a model is the counterpart of Lebesgue measure itself, denoted ${}^*\mathcal{L}$. ${}^*\mathcal{L}$ is defined on the counterpart of the full σ-algebra of Lebesgue-measurable sets and inherits by the Transfer Principle all the first-order properties of $\mathcal{L}$, including translation invariance by any hyperreal. Unfortunately, the Transfer Principle also entails that for each x in ${}^*[0, 1)$, ${}^*\mathcal{L}(\{x\}) = 0$, so that ${}^*\mathcal{L}$ is not regular.

   The adequacy conditions, however, do not require us to use the full ${}^*\mathcal{L}$; perhaps another internal model may be found which is regular (with probability that is merely finitely additive, perhaps, or with a smaller algebra of sets).

   The conditions which make a set function a finitely additive probability on an algebra of subsets of $[0, 1)$ which includes each singleton set can be written down as a sentence of L. In the standard universe, any such set function satisfies (a) $Pr(\{x\}) \geq 0$, and (b) for any finite subset $\{x_1, x_2, \ldots, x_n\} \subset [0, 1)$, $\sum_1^n Pr(\{x\}) \leqslant 1$. It is a familiar fact that no set function satisfying (a) and (b) can be regular; the proof of this is recalled in the appendix. There we prove in fact that any such probability must take the value zero at a dense set of points in $[0, 1)$. This conclusion transfers directly to an internal finitely additive probability on a (nonstandard) algebra of subsets of ${}^*[0, 1)$. Therefore, unfortunately, *no* hyperreal spinner can be regular.

The upshot of this survey of nonstandard models is that we can have a model which is regular, but which is non-translation invariant or external or both, or we can have an internal, translation invariant model which is not regular.

Of course, I have not considered every conceivable nonstandard model which might be proposed as a model for the fair spinner. But the models surveyed thus far pretty well exhaust all obvious candidates for this role. I conclude tentatively that we will not find what we seek by using NSA.

Does Possibility Supervene on Chance?

The negative assessment of all the solutions to the dilemma examined so far might lead us to wonder whether there is any hope of finding a theory in which possibility supervenes on probability. Perhaps we need to take into account 'macroscopic' features of the chance distributions, as well as the probabilities of point outcomes?

One feature of discrete (and especially, of finite) chance distributions that seems to distinguish them from continuous distributions is that in the former, we have little inclination to consider outcomes of chance zero to be possible. For example, in the case of throwing a die, we may with perfect mathematical propriety take the outcome space to be $\{0, 1, 2, 3, 4, 5, 6\}$. It doesn't matter whether we do this because, for example, we incorrectly believe that dice are seven-sided, or because initially it is not settled whether the die comes from a defective batch with one or another numbered side missing, or for some other reason. Later investigation may settle that the die is in fact a normal, symmetric six-sided die. We now have (at least!) two options: we may retain the outcome zero in the outcome space and assign the probabilities $\{0, 1/6, 1/6, 1/6, 1/6, 1/6, 1/6\}$. Or we may excise the outcome zero completely from the outcome space. There is no particular theoretical advantage in pursuing either option; the gain or loss in simplicity is negligible either way.

But the dispensability of the outcome 'zero' in this case renders it implausible to regard this outcome, with its probability of zero, as possible. On the other hand, in the distribution for the fair spinner and in continuous distributions modelling other chance setups, we seem stuck with the apparent possibility of all these outcomes of chance zero. So perhaps the distinction between continuous and discrete distributions could serve as the basis for a theory.

Perhaps, but I doubt it. Consider a modification of the fair spinner, which we may call the Wheel of Fortune. We simply divide the circumference into n "cells" by inscribing lines at $0, 1/n, 2/n, \ldots, (n-1)/n$. We attach a sensor which detects in which cell the pointer stops, and reports the midpoint of that cell as the outcome. For example, if the pointer stops anywhere within $(0, 1/n)$, the sensor outputs the value $1/2n$. But if the pointer stops right on the dividing line between two cells, the sensor, unable to choose between the two cells, simply outputs the position of the dividing line. The sensor forms a filtering 'front end' to the fair spinner, which remains unchanged.

This setup results in a discrete distribution, with outcomes $\{0, 1/2n, 1/n, 3/2n, 2/n, \ldots, (2n-1)/2n\}$. Half the outcomes have chance 0; the other half have chance $1/n$. Are the chance 0 outcomes of the Wheel of Fortune possible? The underlying continuous fair-spinner process, which we modified only by attaching the noninterfering front end, would suggest that they are. But the discreteness of the resulting distribution would suggest that they are not.

Examples such as this, together with the negative results from NSA, cast doubt on the prospects for a theory of chance and possibility in which possibility supervenes on chance. At the least we should not expect a tidy picture in which chance distributions, using the usual possible-world semantics for possibility, can be defined on sets of possible worlds standing in for the outcome spaces. It would seem that neither chance nor possibility furnishes a generally reliable guide to the other.

Appendix: Amplifications and Proofs

Pseudo-Sentences of L In the sentence (Eq. 1) in the text, the symbol $>$ is not an official relation symbol of L. $\mathbf{x} > \mathbf{y}$ is to be regarded as a commonsense abbreviation for a more complicated formula of L, which says that the ordered pair $\langle x, y\rangle$ is a member of the relation "greater-than" which is a subset of $\mathbb{R} \times \mathbb{R}$. If we let the symbol $\mathbf{G}$ name the greater-than relation over $\mathbb{R}$ and $\mathbf{P}$ name the set of (unordered) pairs of elements from $\mathbb{R}$, we can expand $\mathbf{x} > \mathbf{y}$ into a genuine formula of L using bounded quantifiers:

$$(\exists \mathbf{u} \in \mathbf{G})(\exists \mathbf{v} \in \mathbf{P})(\mathbf{x} \in \mathbf{u} \wedge \mathbf{v} \in \mathbf{u} \wedge \mathbf{x} \in \mathbf{v} \wedge \mathbf{y} \in \mathbf{v})$$

(Recall that in set theory, the ordered pair $\langle x, y\rangle$ is the set $\{x, \{x, y\}\}$.)

Abbreviations such as $\mathbf{x} > \mathbf{y}$ are useful in order to make formal sentences intelligible to human readers but do not in principle add any power to L. We might call Eq. 1 a pseudo-sentence, and all such sentences can always with some effort be translated into actual official sentences of L.

L is actually quite a meager language, but it contains the membership symbol $\in$ of set theory. The reader may consult any introductory treatment of axiomatic set theory to see how to build up pseudo-sentences that employ symbols for commonplace functions and relations of mathematics.

External Sets A set is internal if and only if it is a member of a standard set (i.e., a counterpart.) This assures that the set may be the value of some variable in the scope of a bounded quantifier, and thus that it may fall under instances of the Transfer Principle.

Here are some basic results about internal sets, which can be found, for example, in (Lindstrom 1988). Standard sets (counterparts) themselves are always internal. The finite union, intersection, and difference of internal sets are internal. Finite sets of internal sets are internal. A set is internal only if its elements are individuals or internal sets. A function (treated as a set of ordered pairs) is internal only if its domain and range are internal. The inverse of an internal function is internal, and composition of finitely many internal functions is internal.

Countable unions of internal sets are usually *not* internal. For example, the set of counterparts of finite positive integers $\mathbb{N}' = \{{}^*n : n \in \mathbb{N}\}$ is external, even though each singleton $\{{}^*n\}$ is internal, and the superset ${}^*\mathbb{N}$ is internal.

Theorem 1. *$\mathbb{N}'$ is external.*

Proof. Suppose $\mathbb{N}'$ were internal. Since any infinite positive integer M is bigger than each member of $\mathbb{N}'$, it is bounded above. But every *internal* bounded set of real numbers has a least upper bound, by the Transfer Principle. Let B be the least upper bound of $\mathbb{N}'$; B must be infinite. But then $B - 1$ is infinite and also a bound for $\mathbb{N}'$, a contradiction. □

More generally, within the nonstandard universe, *any* infinite set composed purely of standard entities is external. (To be internal, the set has to include nonstandard elements too.) The following Theorem holds:

Theorem 2. *If A is standard and has infinitely many elements, then $A' = \{{}^*a : a \in A\}$ is external.*

Proof. Suppose A' is internal. Let $s = \langle s_1, s_2, \ldots, s_n, \ldots \rangle$ be a standard infinite sequence of distinct elements of A. (In other words, s is a 1–1 function from $\mathbb{N}$ into A.) s has counterpart $^*s = u = \langle u_1, u_2, \ldots, u_n, \ldots, u_M, \ldots \rangle = \langle ^*s_1, ^*s_2, \ldots, ^*s_n, \ldots, u_M, \ldots \rangle$, a 1–1 function from $^*\mathbb{N}$ into $^*A$.

For every *finite* positive integer n, $u_n \in A'$, by the Transfer Principle. I claim that for any *infinite* positive integer M, $u_M \notin A'$. For if $u_M \in A'$, then either $u_M = {^*s_n}$ for some finite positive integer n, or else $u_M = {^*a}$ for some other $a \in A$ which does not occur among the s_n. In the first case, then $(\exists n, m \in {^*\mathbb{N}})(n \neq m \wedge u_n = u_m)$ is true in the nonstandard universe, and hence by transfer $(\exists n, m \in \mathbb{N})(n \neq m \wedge s_n = s_m)$ is true in the standard universe, which contradicts our assumption that s is a sequence of distinct entities. In the second case, then $(\exists n \in {^*\mathbb{N}})(u_n = {^*a})$ is true in the nonstandard universe, and so $(\exists n \in \mathbb{N})(s_n = a)$ is true in the standard universe, which contradicts our assumption that a is not among the s_n.

Since $u_M \notin A'$, we can recover the set of counterparts of standard positive integers thus: $\mathbb{N}' = u^{-1}(A')$. But the right-hand side is internal because u, u^{-1}, and A' are, and that contradicts the externality of $\mathbb{N}'$. □

The Sprinkle Spinner Is Not Translation Invariant This example, derived from an example in Bernstein's and Wattenberg's article, shows that the sprinkle spinner model cannot be translation invariant, even though each real point receives the same infinitesimal measure $\mu(x) = 1/M$, where M is the number of points in the hyperfinite set F. In the following, all additions and multiplications are modulo 1.

Take an irrational $\alpha \in [0, 1)$. Then the set $T = \{\alpha, 2\alpha, 3\alpha, \ldots\}$ consists of countably many distinct points in $[0, 1)$. The points are distinct because if $m\alpha = n\alpha$ for $m \neq n$, then $(m - n)\alpha = k$ for some integer k, contradicting the irrationality of α.

Also, if $T' = T$ translated (shifted rightward) by an amount α, then $T' \subset T$ and in fact $T = \{\alpha\} \bigcup T'$. So the $\mu(T)$ and $\mu(T')$ must differ, by an infinitesimal.

But more is true:

Theorem 3. *We can find two sets X and Y, both of Lebesgue measure zero, which are disjoint and translates of each other, such that $\mu(X) \neq \mu(Y)$ and $\mu(X | X \bigcup Y) \neq 1/2$.*

Proof. Let T and T' be as above. Both are countable and therefore of Lesbesgue measure zero. Choose another irrational β such that β/α is irrational (such a β exists, otherwise there would be only countably many irrationals). Let $X = T + \beta$, i.e. X is T shifted rightward by β. Then X is a countable set and is disjoint from T and from T' (if not, then for some j, k we would have $j\alpha + \beta = k\alpha$, or $\beta = (k - j)\alpha$, contradicting that β/α is irrational). Since $\mu(T) \neq \mu(T')$, either $\mu(X) \neq \mu(T)$ or $\mu(X) \neq \mu(T')$, or both. Take $Y = T$ or $Y = T'$ according to which inequality holds. □

Standard Continuous Probabilities Are Not Regular Theorem 4 is a slightly extended version of a familiar proposition from elementary analysis.

Theorem 4. *Let μ be a finitely additive probability on an algebra of subsets of $[0, 1)$ which includes the singleton $\{x\}$ for each $x \in [0, 1)$. Then μ is not regular: in every subinterval $[a, b] \subset [0, 1)$, there is at least one point y with $\mu(\{y\}) = 0$.*

Proof. Assume μ is regular, and for each positive integer n, define a subset of $[a, b]$ by $S_n = \{x : x \in [a, b] \wedge \mu(\{x\}) > 1/n\}$. Since μ is regular, the union of all the S_n is the whole interval $[a, b]$, which contains uncountably many points, so at least one of the S_n is infinite (a countable union of merely finite sets is still countable).

Let S_n be infinite, and pick distinct points $x_0, x_1, \ldots, x_n$ from S_n. Then by finite additivity, $\mu(\{x_0, x_1, \ldots, x_n\}) \geq (n + 1)/n > 1$, which contradicts that μ is a probability. □

Acknowledgments I am grateful to Ellery Eells and Alan Macdonald for helpful comments on an earlier draft.

Notes

1 We could equivalently use the closed interval [0, 1] with the endpoints identified.

2 In this chapter, I assume that chance is part of the real furniture of the physical universe. Brownian Motion, the random path in space of a small particle immersed in a fluid, may be considered to be a paradigmatic chance process. The status of the fair spinner as a chance device is discussed below.

3 Evidently, the sense of possibility under discussion is not absolute metaphysical or nomological possibility. An alien from Venus might swoop in and demolish the spinner before the pointer stops; but we do not want to recognize "spinner destroyed" as a possible outcome. "Possible" should be construed as "relatively possible," possible relative to fixed conditions which include the physical laws and also initial conditions ensuring that the chance trial terminates successfully.

4 κ-consistency is the generalization of Gödel's ω-consistency to cardinals other than $\aleph_0 = \omega$.

5 Presentation as a limit accounts for the overwhelming importance of the normal distribution in statistical theories, since the Central Limit Theorem assures us that a host of different sequences of distributions will converge weakly to the normal distribution. Two examples important in practice are sequences of (suitably normalized) binomial and poisson distributions. The uniform distribution also may be multiply presented in this way, as the limit of different sequences of distributions, with different properties.

6 Here I borrow from Lewis a terminology which seems apt, but which he used for a completely different purpose.

7 In a bounded sentence, all the quantifiers have the form $\exists \mathbf{x} \in \mathbf{s}$ or $\forall \mathbf{x} \in \mathbf{s}$, where $\mathbf{s}$ names a standard entity, or else $\exists \mathbf{x} \in \mathbf{y}$ or $\forall \mathbf{x} \in \mathbf{y}$, where $\mathbf{y}$ is a variable. The effect is to restrict the domain of discourse of a sentence to the elements of some standard entity or counterpart (or the elements of an element, etc.).

8 This sentence is not actually in the language L, because of the use of the symbol $\mathbf{>}$. This turns out not to be a real problem; see the appendix for a discussion.

9 External sets can play a role in relating purely nonstandard results back to the standard universe. For example, the "standard part" mapping, which maps a limited (i.e., non-infinite) hyperreal back to the unique standard real which is infinitely close to it, is an external mapping (viewed as a set of ordered pairs, it is an external set).

10 Finite probability theory has the instrumental virtue of being comparatively easy conceptually and computationally. Nelson (1987) has given a readable exposition of a big part of probability theory reconstructed with hyperfinite sets.

11 Although we cannot straightforwardly say that the lack of translation invariance is *because* of the externality, the fact remains that this model is deficient in deducible properties useful for calculation and prediction, and that it is reasonable to expect other external models to suffer from the same problem.

References

Bartha P, Johns R (2001) Probability and symmetry. Philos Sci 68 (Proceedings)

Bernstein RA, Wattenberg F (1969) Nonstandard measure theory. In: Luxemburg WAJ (ed) Applications of model theory to algebra, analysis, and probability. Holt, Rinehart, and Winston, New York, pp. 171–185

Elga A (Mar 2004) Infinitesimal chances and the laws of nature. Aust J Philos 82(1):67–76

Lewis D (1973) Causation. J Philos 70:556–567

Lindstrom T (1988) An invitation to nonstandard analysis. In: Nigel C (ed) Nonstandard analysis and its applications, number 10 in London Mathematical Society Student Texts. Cambridge University Press, Cambridge

McGee V (1994) Learning the impossible. In: Eells E, Brian S (eds) Probability and conditionals. Cambridge University Press, Cambridge

Nelson E (1987) Radically elementary probability theory. Number 117 in annals of mathematics studies. Princeton University Press, Princeton

Nelson E (2001) Internal set theory, 2001. http://www.math.princeton.edu/ nelson/books.html

Ross D (1997) Loeb measure and probability. In: Nigel Cutland Leif O. Arkeryd, Ward Henson C (eds) Nonstandard analysis, theory and applications. Kluwer, Dordrecht, pp 91–120

Skyrms B (1980) Causal necessity. Yale University Press, New Haven

Schechter E (1997) Handbook of analysis and its foundations. Academic, San Diego

van Fraassen B (1980) The scientific image. Oxford University Press, Oxford

Zaman A (1987) On the impossibility of events of measure zero. Theor Decis 23:157–159

Probabilistic Metaphysics

James H. Fetzer

The demise of deterministic theories and the rise of indeterministic theories clearly qualifies as the most striking feature of the history of science since Newton, just as the demise of teleological explanations and the rise of mechanistic explanations dominates the history of science before Newton's time. In spite of the increasing prominence of probabilistic conceptions in physics, in chemistry and in biology, for example, the comprehensive reconciliation of mechanistic explanations with indeterministic theories has not gone smoothly, especially by virtue of a traditional tendency to associate "causation" with determinism and "indeterminism" with non-causation. From this point of view, the very idea of *indeterministic causation* seems to be conceptually anomalous if not semantically inconsistent. Indeterminism, however, should not be viewed as the absence of causation but as the presence of causal processes of non-deterministic kinds, where an absence of causation can be called "non-causation".

The underlying difference between causal processes of these two kinds may be drawn relative to events of one type (radioactive decay, genic transmission, coin tossing, and so on) as "causes" C of events of other types (beta-particle emission, male-sex inheritance, coming-up heads, and so forth) as "effects" E. Then if deterministic causal processes are present whenever "the same cause" C invariably brings about "the same effect" E, then indeterministic causal processes are similarly present whenever "the same cause" C variably brings about "different effects" $E^1, E^2, \ldots$ belonging to some specific class of possible results. So long as beta-particle emission (male-sex inheritance, coming-up heads, ...) are invariable "effects" of radioactive decay (of genic transmission, of coin tossing, ...), respectively, then these are "deterministic" causal processes; but if alpha-particle emission (female-sex inheritance, coming-up tails, ...) are also possible, under the very same conditions, respectively, the corresponding causal processes are "indeterministic", instead.

A deterministic theory (or a deterministic law), therefore, characterizes every physical system of a specified kind K as an instance of a deterministic causal process for which "the same cause" C invariably brings about "the same effect"

J.H. Fetzer (✉)
McKnight Professor Emeritus, University of Minnesota Duluth, 800 Violet Lane, WI 53575, Oregon
e-mail: jfetzer@d.umn.edu

E. Eells and J.H. Fetzer (eds.), *The Place of Probability in Science*, Boston Studies in the Philosophy of Science 284, DOI 10.1007/978-90-481-3615-5_4,

E, where indeterministic theories (and indeterministic laws) are open to parallel definition. Thus, the world itself **W** is an indeterministic system if at least one indeterministic theory describing the world is true, which will be the case if at least one of its causal processes is indeterministic in kind. The conceptual problem that remains, of course, is understanding how it can be possible, in principle, for "different effects" E^1, E^2, ... to be brought about by "the same cause" C, i.e., under the very same conditions. Recent work within this area, however, suggests the plausibility of at least two different approaches toward understanding "indeterministic causation", namely: various attempts to analyse "causation" in terms of probability, on the one hand, and various attempts to analyse "probability" in terms of causation, on the other.

Since attempts to analyse "causation" in terms of probability tend to be based upon interpretations of probabilities as actual or as hypothetical limiting frequencies, moreover, while attempts to analyse "probability" in terms of causation tend to be founded upon interpretations of probabilities as long-run or as single-case propensities instead, let us refer to frequency-based accounts as theories of "statistical causality" and to propensity-based accounts as theories of "probabilistic causation". Even though there thus appear to be two different types of each of these approaches toward this critical conceptual problem, it is not obvious that any of these theories can satisfy the relevant desiderata. My purpose here is to review why three out of four hold no promise in the reconciliation of mechanistic explanations with indeterministic theories. If these reflections are well-founded, then the single-case propensity approach alone provides an appropriate conception of the causal structure of the world.

Relevant Desiderata: Conditions of Adequacy

The first task confronting this investigation, therefore, is the specification of suitable desiderata for an adequate construction of indeterministic causation. Some of these criteria are required of "probabilistic" interpretations, while others are imposed by the element of "causality". We may assume, as probabilistic conditions, the requirements of admissibility, of ascertainability, and of applicability advanced by Salmon (1967), given the following understanding: (1) that the admissibility criterion requires that relations characteristic of mathematical probabilities (such as those of addition, of summation, and of multiplication) must be satisfied rather than some specific formulation of the calculus of probability; (2) that the ascertainability criterion requires that methods appropriate to statistical inquiries (such as statistical hypothesis-testing techniques) are capable of subjecting probability hypotheses to empirical tests but not that they should be verifiable; and, (3) that the applicability criterion requires not only that probabilities have to be predictively significant for "the long run", but also for "the short run" and for "the single case", i.e., for infinite, finite and singular sequences.

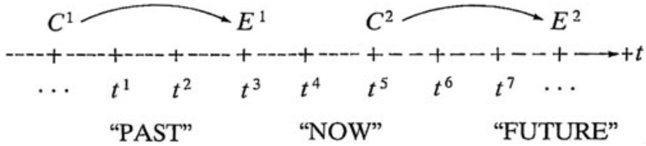
These conditions are important for a variety of reasons. In the first place, if condition (1) required that an acceptable interpretation must satisfy some specific formulation of the calculus of probability rather than relations characteristic of mathematical probabilities, propensity conceptions – which cannot satisfy Bayes' theorem, for example – would be excluded a priori. In the second place, if condition (2) required that an acceptable interpretation must be verifiable rather than merely empirically testable, only finite relative frequency constructions – but no limiting frequency or any propensity conceptions – could possibly qualify as acceptable. Thus, since condition (3) appears to impose a requirement that only single-case conceptions could be expected to fulfill, further consideration should be given to its justification and to the possibility that, in principle, it ought to be weakened or revised.

In order to fulfill the function of reconciling mechanistic explanations with indeterministic theories, furthermore, an acceptable account of indeterministic causation ought to generate "mechanistic explanations" that explain the present and the future in terms of the past, rather than conversely. Indeed, this condition accords with several features that seem to be characteristic if not invariable aspects of deterministic causation, such as (a) that "causes" precede their "effects", (b) that "effects" do not bring about "causes", and (c) that "causes" differ from their "effects". Let us also assume that "teleological explanations" either (i) explain the past and the present in terms of the future (the *temporal* criterion) or (ii) explain "effects" by citing motives, purposes, or goals (the *intentional* criterion). Thus, some "mechanistic explanations", none of which are "teleological$_1$" in the strong temporal sense, might still be "teleological$_2$" in the weak intentional sense, which we exemplify when we explain our behavior by citing motives and beliefs.

While "intentional explanations" need not explain the past and the present in terms of the future, therefore, an acceptable account of indeterministic causation should not generate "teleological explanations" of either kind, except with respect to appropriate classes of instances involving the proper ascription of *propositional attitudes* as they occur, for example, within psychology, sociology, and anthropology. Ascribing motives, purposes, or goals to inanimate objects, such as gambling devices and subatomic particles within physics or chemistry, appears to entail the adoption of animistic hypotheses, which certainly should not be a pre-condition for "indeterministic causation". And, insofar as causation assumes its fundamental, albeit not exclusive, role with respect to explanation, we may also suppose, as conditions of causation, that an adequate conception of this kind ought to possess explanatory significance not only for "the long run" but also for "the short run" and for "the single case". Though these conditions jointly entail that an acceptable conception of indeterministic causation should support probabilistic hypotheses that are both predictively and explanatorily significant for infinite, finite and singular sequences, they do not dictate any particular variety of "causal explanations", provided they are not "teleological explanations", necessarily.

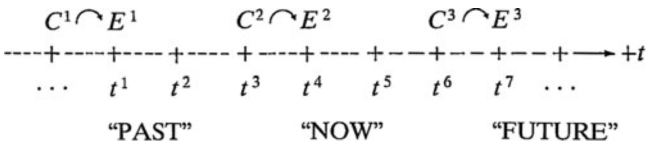
Strictly speaking, of course, it is suggestive but inaccurate to say that teleological explanations explain the past and the present in terms of the future, since the temporal criterion hinges upon relations of "earlier than" and of "later than". Indeed, four more-or-less distinct species of causation may be identified by introducing "proximate" and "distant" differentia as follows:

(A) *Distant mechanistic causation*



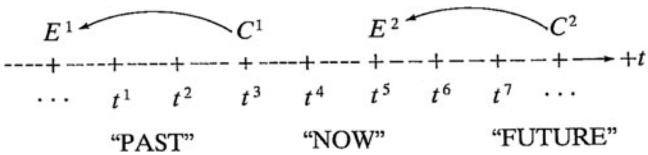
in which spatio-temporally separated "causes" C^1, C^2, ... bring about their "effects" E^1, E^2, ... (roughly) by forward "causal leaps" within space-time – whether these causal leaps belong to the "PAST", the "NOW", or the "FUTURE";

(B) *Proximate mechanistic causation*



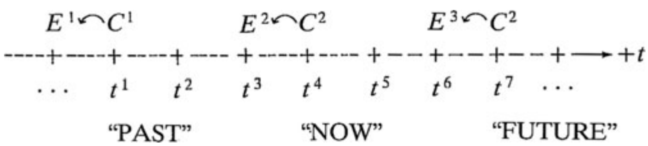
in which spatio-temporally contiguous "causes" C^1, C^2, ... bring about their "effects" E^1, E^2, ... by (minimal) forward "causal steps" within space-time – whether these causal steps belong to the "PAST", the "NOW", or the "FUTURE";

(C) *Distant* teleological$_1$ *causation*



in which spatio-temporally separated "causes" C^1, C^2, ... bring about their "effects" E^1, E^2, ... by backward "causal leaps" within space-time – whether these causal leaps belong to the "PAST", the "NOW", or the "FUTURE"; and, finally,

(D) *Proximate teleological$_1$ causation*



in which spatio-temporally contiguous "causes" C^1, C^2, ... bring about their "effects" E^1, E^2, ... by (minimal) backward "causal steps" within space-time – whether these causal steps belong to the "PAST", the "NOW", or the "FUTURE".

While "teleological causation" (in the sense of teleology$_1$) characteristically requires backward (or "retro-") causation, "distant causation" typically entails causation-at-a-distance (within space-time), contrary to the principle of locality of special relativity, which postulates the following requirement:

(E) *The principle of locality* causal influences are transmitted by means of causal processes as continuous sequences of events with a finite upper bound equal to the speed of light ($\approx$186,000 mps)

hence, unless special relativity is to be rejected, there are spatio-temporal restrictions upon causal processes that render both varieties of "distant causation" unacceptable, in principle, since "causal leaps" of either kind would entail violations of the principle of locality. Thus, classical behaviorism, with its tendency to assume that an individual's past history directly determines his present behavior (as an example of distant mechanistic causation), is no more acceptable with respect to its explanatory type within psychology than are appeals to "manifest destiny", with their tendency to presume that present actions are directly determined by future events (as illustrations of distant teleological$_1$ causation), in their explanatory roles within history.

The principle of locality, of course, may be reformulated to reflect the (discrete or continuous) character of space-time by substituting "contiguous" for "continuous" in sentence (E), just as diagrams (A) through (D) may be employed to represent (deterministic or indeterministic) causal processes with the utilization of continuous and discontinuous "causal arrows", respectively. Hence, indeterministic as well as deterministic varieties of distant and proximate (teleological and mechanistic) causation are logical possibilities that do not necessarily also qualify as physical possibilities. Considerations of these kinds, moreover, have very significant consequences for alternative interpretations of indeterministic causation. For, with respect to "the single case", "the short run" and "the long run", conceptions of statistical causality based upon actual and hypothetical limiting frequencies result in probabilistic hypotheses that either are purely descriptive and non-explanatory or else exemplify distant teleological$_1$ causation, while conceptions of probabilistic causation based upon long-run propensities, by comparison, result in probabilistic hypotheses that exemplify both distant mechanistic and proximate teleological$_2$ species of causation, as the arguments that now follow are intended to display.

Statistical Causality: Actual Limiting Frequencies

Let us begin with theories of statistical causality based upon the actual limiting frequency construction of objective probabilities, which, in turn, is built upon (a) the mathematical definition of the limit of a relative frequency within an infinite sequence, (b) the empirical interpretation of such limits as features of the actual world's own history, and (c) the logical formulation of these probabilistic properties as conditional probabilities (cf. Salmon 1967). Thus, under this construction, the probability of an event of a specific type, Y, such as beta-particle emission, male-sex inheritance, etc., relative to the occurrence of an event of another specific type, X, such as radioactive decay, genic transmission, etc., can be characterized, in general, along these lines:

(F) $P(Y/X) = p =_{\mathrm{df|}} \lim n \to \infty f^n(Ym/Xn) = (p = m/n)$

that is, the probability for (an event of kind) Y, given (an event of kind) X, has the value p if and only if (events of kind) Y occur relative to (events of kind) X with a limiting frequency m/n equal to p during the course of the actual history of the world. Insofar as limiting frequencies for events of particular kinds may vary relative to different reference classes $X,\ X',\ X''$ and so forth, various relations of statistical relevance can be defined, such as, for example,

(G) If $P(Y/X \ \&\ Z) \neq P(Y/X \ \&\ -Z)$, then the occurrence of (an event of kind) Z is statistically relevant to the occurrence of (an event of kind) Y, relative to the occurrence of (events of kind) X

which Salmon (1970) utilized as the foundation for a new account of statistical explanation intended to improve upon Hempel's inductive-statistical theory (Hempel 1962, 1965, 1968). While the mathematical definition of limiting frequency is straightforward when applied within abstract domains (such as the theory of numbers), however, its introduction within causal contexts has raised a variety of problems, most of which are now familiar but remain important, nevertheless.

One such difficulty arises because limiting frequencies, as properties of infinite sequences, are well-defined as features of the physical world only if the number n of occurrences of reference events (of kind X, say) increases without bound. This means that if there were no – or only a few, or many but finite – instances of events of kind X, then neither the probability for (an event of kind) Y given (an event of kind) X nor the relevance relations between (events of kinds) X, Y and Z (for events of any other kind) would be well-defined. So unless there is no end to the number of instances of genic transmission (radioactive decay, etc.) during the history of the actual world **W**, there can be no corresponding limiting frequencies to satisfy the conditions specified by (F) and by (G). Moreover, as long as these events themselves are non-vanishing in their duration, such a conception also entails that either the world's history must go on forever or else the corresponding probabilities cannot exist. So if the existence of limiting frequencies is required for "statistical causality", then neither (probabilistic) "causes" nor "effects" can exist if the world's history is finite.

This problem has invited a number of attempted solutions, the majority of which, following von Mises (1964), treat the infinite sequence condition as an "idealization" concerning what the limiting frequencies would be (or would have been) if the world's history were (or had been) infinite. This defense itself, however, appears to raise more questions than it answers, especially regarding the theoretical justification for these subjunctive and counterfactual claims. An ontological warrant might be generated from the ascription of dispositional properties to the physical world (as an inference to the best explanation, for example), where these properties could be invoked as a foundation for predicting and explaining the occurrence of relative frequencies; but this, as later discussion of the hypothetical frequency and of alternative propensity conceptions should make evident, would be to abandon rather than to repair this position.

Alternatively, a psychological warrant for ascribing hypothetical properties could be discerned in our (inevitable) "habits of mind" within the framework pioneered by Hume; yet, even if our anticipatory tendencies were thereby accurately described, that in itself would provide no foundation for their justification. After all, that we have such expectations does not establish that we ought to have them (although an evolutionary argument could be made to that effect, which would presumably bear the burden not only of separating adaptive from maladaptive expectations but also of deriving their truth from their benefits to survival and reproduction). The evolutionary benefits of expectations almost certainly derive from those "habits of mind" that obtain with respect to the single case and the short run, however, whose precise relations to the long run are not entirely clear. Moreover, while it may be plausible to reason from the truth of an expectation to its (potential) evolutionary benefits, to reason from its (potential) evolutionary benefits to its truth is to beg the question.

Other difficulties arise insofar as limiting frequencies, as symmetrical properties, violate our assumptions about causal relations. Consider, for example, that if 'X' and 'Y' are interpreted as "cause" and "effect", then since

(H) $P(X \,\&\, Y)/P(X) = P(X/Y)P(Y)/P(X)$
$= P(Y \,\&\, X)/\, P(X) = P(Y/X)P(X)/P(X)$

– where common reference classes have been suppressed for convenience – whenever "causes" X bring about "effects" Y, those "effects" Y must likewise bring about those "causes" X, in violation of the presumption that "effects" do not (simultaneously) bring about their "causes". Moreover, if 'X' and 'Y' are interpreted as singular events '*Xat*' and '*Yat**', where $t^* > t$, rather than as event types,

(I) $P(Yat^*/Xat)P(Xat)/P(Xat) = P(Xat/Yat^*)P(Yat^*)/P(Xat)$

then, whenever *Xat* "causes" *Yat**, *Yat** must also "cause" *Xat*, in violation of the presumption that "causes" precede their "effects". One avenue of escape from criticisms of this kind, however, might be afforded by emphasizing the temporal aspect of causal relations, where only conditioning events, such as *Xat*, that are earlier than conditioned events, such as *Yat**, can possibly qualify as "causes" in relation to "effects". Such a maneuver entails the result that, with respect to formulae like (H) and (I), left- and right-hand equalities in numerical value need not reflect similarities in semantical meaning.

Difficulties such as these, of course, have not inhibited Good (1961/62), Suppes (1970), and especially Salmon (1975, 1978, 1980) from elaborating frequency-based accounts of "statistical causality" [although, to be sure, more recently Salmon has endorsed a propensity approach; cf. esp. Salmon (1984)]. Not the least of the problems confronting this program, however, are (i) that statistical relevance relations – positive, negative or otherwise – do not need to reflect causal relevance relations (even when temporal conditions are introduced); and, (ii) that the systematic application of frequency-based criteria of causal relevance (such as those of Suppes and of Salmon) yields the consequence that particular attributes within non-epistemic homogeneous partitions of reference classes invariably occur only with degenerate probabilities equal to zero or equal to one. If statistical relevance relations and temporal conditions were enough to define causal relevance relations, it would turn out to be logically impossible, in principle, for causation to be indeterministic.

As an illustration of this situation (which does not require elaborating these arguments in great detail), let us consider the analysis of statistical causality advanced by Suppes (1970), where he suggests that the basic concept is that of "*prima facie* causation" where X is a *prima facie* "cause" of Y when and only when (a) the event X occurs earlier than the event Y, i.e., if *Yat**, then *Xat*, where $t^* > t$; and, (b) the conditional probability of Y given X is greater than the unconditional probability of Y, i.e., $P(Y/X) > P(Y)$. Notice that positive statistical relevance is employed here as a measure of positive causal relevance, i.e., Suppes assumes that conditions in relation to which Y occurs more frequently are causally relevant conditions, so that if lung cancer occurs with a higher frequency in smoking populations than in non-smoking populations, then smoking is a *prima facie* cause of lung cancer. To complete his analysis, Suppes further defines a "genuine" cause as a *prima facie* cause that is not a "spurious" cause, which requires that there be no earlier event Z such that the conditional probability of Y given Z & X equals that of Y given X, i.e., $P(Y/Z \,\&\, X) \neq P(Y/X)$. Accordingly, the "genuine" cause of an event Y is the event X earliest in time relative to which the probability of Y is highest.

In other to appreciate the magnitude of the obstacles confronting Suppes' constructions, keep in mind that any instance of lung cancer Y will involve a specific person i who possesses many properties in addition to being a smoker or not, as the case happens to be; for any such person will also be married or single, drink gin or abstain, eat caviar or ignore it. Indeed, the frequency for lung cancer Y (or for its absence $-Y$) will vary from class to class as additional properties $F^1, F^2, \ldots$ are taken into account until the upper bound of one has been reached; for if i is a gin-drinking, caviar-eating, heavy-smoking married man, there must exist some reference class, X^*, to which i belongs and in relation to which Y (or $-Y$) occurs more frequently than it occurs in relation to any other reference class to which i belongs, namely: a homogenous class X^* for which every member of X^* has lung cancer Y (or has no lung cancer $-Y$), as it turns out; otherwise, Suppes' conditions cannot have been satisfied. For such a reference class, however, Y (or $-Y$) occurs with probability of one; hence, on Suppes' account, "statistical causality" is a logical impossibility.

These difficulties, of course, are not really hard to understand, insofar as (i) limiting frequencies, as a species of statistical correlation, are only (at best) necessary but not sufficient conditions for causal relations; while, (ii) as properties of the members of reference classes collectively, limiting frequencies are not properties of each member of these classes individually. Consequently, analyses of "statistical causality" that are based upon actual limiting frequencies, as "long run" distributions during the world's history, are purely descriptive and non-explanatory. Even in conjunction with temporal conditions, they are incapable of satisfying the requirements for causal relations. Moreover, as properties of infinite event sequences, they cannot even qualify as properties of singular events themselves. Indeed, although Salmon (1975) thought that causal relevance relations might be analysed as (complex) statistical relevance relations, his more recent efforts (Salmon 1978, 1980, 1984) suggest, regarding statistical causality, that an analysis in terms of statistical relevance relations and *independent* causal concepts invites a regress, while an analysis in terms of statistical relevance relations and *dependent* causal concepts cannot be correct. Indeed, from premises confined to correlations, only conclusions concerning correlations can be validly derived, and they exemplify neither mechanistic nor teleological species of causation.

Statistical Causality: Hypothetical Limiting Frequencies

If these considerations are well-founded, then an analysis of statistical causality based upon the actual limiting frequency construction offers no hope for understanding indeterministic causation; thus, perhaps an account based upon the hypothetical limiting frequency construction might fare a little better. Now the principal difference between the actual and the hypothetical frequency conceptions is not a matter of mathematical frameworks but rather of empirical interpretation, since these limits are features, not of the actual history of the world, **H**(**W**), but of hypothetical extensions of that history, $H_i(\mathbf{W})$. As a consequence, these objective probabilities may be identified with the limiting frequency, if any, with which Y would occur, were the number of events of kind X that occur during the history of the world to increase without bound (as an illustration, see van Fraassen 1980). The hypothetical frequency approach thus promises to improve upon the actual frequency approach in at least three respects, since (a) hypothetical sequences are available, even when actual sequences are not, ensuring that relevant reference classes have enough members; (b) systematic comparisons of statistical correlations then may be refined for improved analyses of causal relations; and, (c) specific attributes within nonepistemic homogeneous partitions of these reference classes perhaps might occur with non-degenerate probabilities. As appealing as this account may initially appear, however, it confronts several rather imposing difficulties of its own.

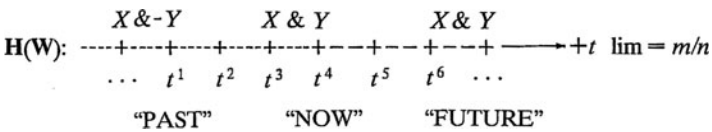
The major difficulty arises from the hypothetical character of these "long run" frequencies, a linguistic manifestation of which is the use of subjunctive conditionals (concerning what limits would obtain if sequences were infinite) in lieu of indicative conditionals (concerning what limits do obtain when sequences are infinite). Thus, within scientific discourse, at least, there appear to be only two modes of justification for subjunctive assertions, namely: (i) those that can be sustained on *logical* grounds alone (as a function of meaning or of grammar), such as, "If John were a bachelor, then John would be unmarried", on the one hand, and (ii) those that can be sustained on *ontological* grounds (as a function of causal or of nomic necessity), such as, "If this were made of gold, then it would melt at 1063°C", on the other. Frequentist analyses of statistical causality, however, are unable to secure support from either direction, for limiting frequencies, whether actual or hypothetical, are supposed to be logically contingent; and, as frequency distributions, with or without limits, they are supposed to forego commitments to the existence of non-logical necessities.

From an epistemological point of view, the actual limiting frequency construction seems to fare far better, since actual limiting frequencies, as features of the world's actual history, either occur as elements of that history or they do not, while hypothetical frequency hypotheses, as designating extensions of the world's actual history, are supposed to be true or false whether or not these occur as elements of that history. From an ontological point of view, moreover, actual limiting frequencies are purely descriptive and non-explanatory because they merely describe statistical correlations that happen to occur during the history of the world, which means that statements describing them are *extensional generalizations* as truth-functions of the truth-values of enormously numerous, purely descriptive singular sentences, in turn. Insofar as hypothetical limiting frequencies do not merely describe statistical correlations that happen to occur during the history of the world, however, statements describing them cannot be truth-functions of the truth-values of (even) enormously numerous, purely descriptive singular sentences, which means that, under the hypothetical frequency construction, these probabilistic hypotheses have to be *non-extensional generalizations* that are no longer truth-functional.

Although the hypothetical frequency approach provides *prima facie* support for subjunctive conditionals (concerning what limits would obtain if sequences were infinite, whether or not they actually are) and for counterfactual conditionals (concerning what limits would have obtained if sequences had been infinite, when they actually are not), therefore, the problem still remains of explaining why these statements – characterizing what these limiting frequencies would be or would have been if these sequences were or had been infinite – are true (cf. Fetzer 1974). In contrast with the actual frequency approach, which does not require them, in other words, the hypothetical frequency construction provides no theoretical principles or structural properties which might be invoked to explain the attribution of these hypothetical limiting frequencies to the physical world. This approach thus appears to be incapable, in principle, of affording a theoretical justification for its own probabilistic hypotheses – unless these limits are viewed as aspects of the world's "manifest destiny".

Indeed, there are several reasons to believe that any analysis of statistical causality that is based upon the hypothetical frequency interpretation will be essentially – and not merely incidentally – committed to distant teleological$_1$ causation. One such reason arises from the theoretical necessity to establish ordering relations within these classes of hypothetical events, since limiting frequencies are subject to variation with variations in ordering relations. An analysis based upon the actual frequency interpretation, by contrast, does not encounter this problem, because temporal relations between actual events serve to fix their order, which is a feature of the actual history **H** of the world **W**:

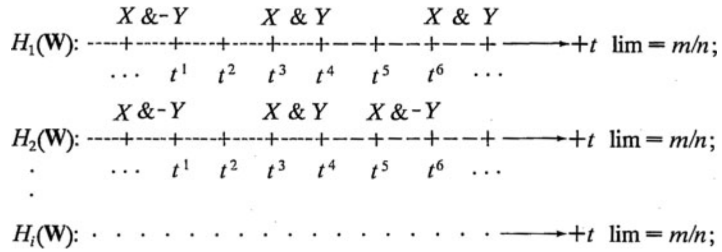
(J) *Statistical causality as a function of actual frequencies*



where m/n represents the conditional probability for Y given X as the limiting frequency with which Y-events occur relative to X-events during the world's history.

Although a single infinite sequence is sufficient to exhibit the actual history of the world, an infinite sequence of infinite sequences is required to exhibit the enormous range of hypothetical extensions of the world's history up to the present, signified by "NOW", where, to guarantee these limits are unique, the same limits must obtain no matter how the world's history might "work out", i.e., the same attributes must occur with the same limits in each such history:

(K) *Statistical causality as a function of hypothetical frequencies*



thus, no matter what the specific features of those singular histories may be, the same attributes must occur with the same limiting frequency m/n in each of those hypothetical extensions of the world's actual history, $H_i(\mathbf{W})$. Moreover, it is not difficult to establish that there must be an endless number of such hypothetical extensions by a diagonal argument, since, for any specific class thereof, simply construct a new sequence that differs from the first sequence in its first hypothetical "segment", from the second in its second, and so on.

If this conception guarantees that these limits are unique, it has to pay the price; for, since the hypothetical frequency approach provides no mechanistic properties, teleological$_2$ or not, that might be invoked to explain the ascription of these hypothetical extensions of the world's actual history, there appears to be no alternative explanation than to conclude that they occur as a reflection of the world's "manifest destiny", i.e., as manifestations of distant teleological$_1$ causation. Indeed, if these limiting frequencies themselves seem to reflect distant teleological$_1$ causation, consider how much worse things are for "the short run" and for "the single case", since finite frequencies and singular events are only "explainable", within the framework of this conception, through their assimilation as individually insignificant, incidental features of a single "grand design" whereby the future brings about the present and the past. Even without the added weight of the symmetrical properties that accompany any attempt to envision indeterministic causation as a species of conditional probability, these are grounds enough to embrace the conclusion that hypothetical frequency based theories of statistical causality are very unlikely to succeed.

Probabilistic Causation: Long-Run Propensities

The failure of frequency-based accounts of indeterministic causation as theories of statistical causality, no doubt, strongly suggests that alternative accounts must be considered, including, in particular, propensity-based analyses of "probabilistic causation". Propensity conceptions of objective probabilities as dispositional properties are built upon (a) the interpretation of dispositions as causal tendencies (to bring about specific outcomes under relevant test conditions), (b) where these causal tendencies are properties of (instantiations of) "probabilistic" experimental arrangements (or "chance set-ups"), (c) which are subject to formalization by "probabilistic" causal conditionals incorporating a primitive *brings about* relation (Fetzer 1981, Ch. 3). The most important difference between theories of statistical causality and theories of probabilistic causation is that, on propensity-based theories, probabilities *cause* frequencies, while on frequency-based theories, probabilities *are* frequencies. The long-run propensity construction thus identifies probabilities with "long-run" causal tendencies, while the single-case propensity construction identifies them with "single-case" causal tendencies, for relevant "trials" to bring about appropriate "outcomes".

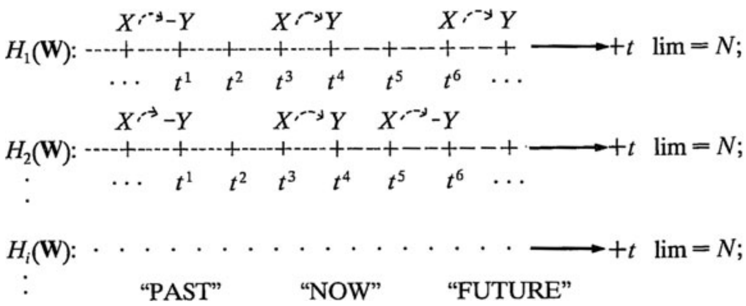
The objective probability of an event of a particular type, Y (such as beta-particle emission, male-sex inheritance, . . .), therefore, is either identified with a dispositional tendency (of universal strength) for events of kind Y to be brought about with an invariable limiting frequency or with a dispositional tendency (of probabilistic strength) for events of kind Y to be brought about with a constant single-case propensity. The most important difference between the hypothetical frequency conception and these propensity conceptions, furthermore, is that long-run and single-case propensities, as long-run and single-case causal tendencies, qualify as mechanistic properties that might be invoked to explain the ascription of hypothetical extensions of the world's actual history, since frequencies are brought about by propensities. Because these properties are formalized as asymmetrical causal tendencies for events of kind Y to be brought about by events of kind X with propensity N,

(L) $Xat \ni\!\!\rightarrow_N Yat^*$

rather than as symmetrical conditional probabilities, it is not the case that whenever Xat "causes" Yat^*, Yat^* must also "cause" Xat, thereby conforming to our expectations about causal relations. In spite of these similarities, however, these propensity conceptions are dissimilar in several crucial respects.

The long-run propensity conception identifies probabilities with what the limiting frequencies for Y are or would be or would have been, had an appropriate chance set-up been subject to a suitable (infinite) sequence of trials, X, which may be characterized in relation to possible histories H_i of the world $\mathbf{W}$:

(M) *Probabilistic causation as a function of long-run propensities*



where, no matter what specific history the world's actual history **H** might have displayed or might yet display, the same specific "effects" Y, $-Y$, ..., are invariably brought about with limiting frequencies N equal to their generating "long-run" propensities, whenever those histories reflect an infinite sequence of trials of the relevant X-kind. As a consequence, the "long-run" propensity conception ascribes a causal tendency of universal strength to bring about outcomes of kind Y with (invariable) limiting frequency N to every "chance set-up" possessing that dispositional property if subject to a trial of kind X, which, in the case at hand, itself consists of an infinite class of singular X-trials.

Since the "causes" X of these limiting frequencies for their "effects" Y themselves turn out to be infinite classes of spatially distributed and temporally extended singular trials, where "long run" dispositional tendencies serve as mechanistic properties, it should come as no surprise that this "long run" account exemplifies distant mechanistic causation (in relation to these infinite classes of trials). It is even more intriguing, therefore, to contemplate the contribution that each singular trial, *Xat*, must make to the attainment of that ultimate "effect"; for, unless each of these singular trials, so to speak, is able to "keep track" of the relative frequencies for outcomes of kind Y, $-Y$, ..., it would be impossible to guarantee the same outcomes invariably occur with the same limiting frequencies during every possible history of the world. And if that, indeed, is the case, then, in effect, each singular trial member of these infinite trial sequences must not only "keep track" of how things are going but also "act" with the intention of insuring that things work out right, which means that this account also exemplifies proximate teleological$_2$ causation (in relation to each such singular trial). But surely no analysis of indeterministic causation incorporating both distant mechanistic and proximate teleological$_2$ species of causation could possibly satisfy the appropriate desiderata.

Probabilistic Causation: Single-Case Propensities

If an analysis of probabilistic causation can attain the reconciliation of mechanistic explanations with indeterministic theories, therefore, it must be based upon the single-case propensity conception. The most important difference between "long run" and "single case" propensities is that "single case" propensities are single-case as opposed to long-run causal tendencies, where their "effects" are the (variable) results (such as coming up heads or coming up tails, ...) of singular trials (such as a single toss of a coin, ...) rather than any (invariable) outcomes (in the form of limiting frequencies) of infinite sequences of trials (Fetzer 1971). Single-case propensities, as properties of chance set-ups, also qualify as mechanistic properties that might be invoked to explain ascriptions of hypothetical extensions of the world's actual history – not collectively (for infinite classes of trials), but distributively (for each specific individual trial), where collective consequences for finite and for infinite sets of trials follow by probabilistic calculations. Single case propensities, moreover, bring about specific outcomes Y, $-Y$, ..., with constant (probabilistic) strength from trial to trial, but they generate only variable relative and limiting frequencies over finite and infinite sequences of trials, unlike any of the other interpretations we have considered before.

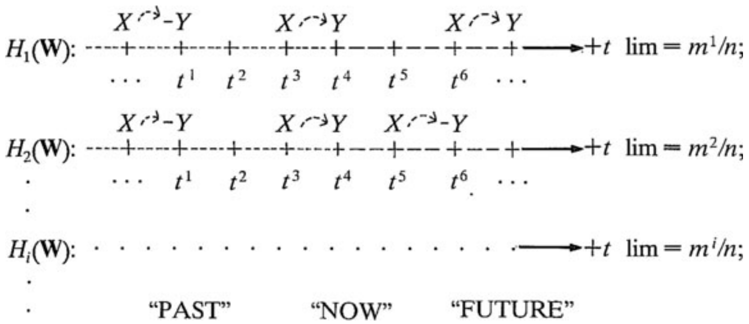
As we have already discovered, frequency-based constructions of statistical causality can be employed to generate frequency-based criteria of statistical relevance, such as (G), which can be utilized as the foundation for frequency-based accounts of statistical explanation (cf. especially Salmon 1971). Analogously, propensity-based conceptions of probabilistic causation can be employed to generate propensity-based criteria of causal relevance, such as:

(N) If $[(Xat\ \&\ Zat) \ni_N Yat^*]$ and $[(Xat\ \&\ -Zat) \ni_M Yat^*]$, where $N \neq M$, (the property) Z is causally relevant to the occurrence of (the property) Y, in relation to the occurrence of (the property) X

which can be employed as the foundation for propensity-based analyses of probabilistic explanation (cf. Fetzer 1981, Part II). Thus, while frequency-based criteria entail the consequence that statistically-relevant properties are therefore explanatorily-relevant properties, propensity-based criteria entail the consequence that statistically-relevant properties may or may not be causally-relevant or explanatorily-relevant as well. Indeed, the most striking development in current work within this field has been the abandonment of a frequency-based model of statistical explanation in favor of a propensity-based model of causal explanation by Salmon (1984). (The extent to which Salmon has carried out this crucial conceptual exchange receives consideration in Fetzer (1987)).

The single-case propensity construction does not identify "probabilities" with what the limiting frequencies are or would be or would have been, therefore, because, in relation to the single-case conception, arbitrary deviations between the strength of a probabilistic "cause" and the limiting frequency of its "effects", even over the long run, are not merely logically possible, but nomologically expectable, where expectations of deviations are open to systematic computation. Unlike the hypothetical frequency and the "long run" propensity conceptions, in other words, the causal consequences of "single case" propensities now may be accurately represented by an infinite collection of infinite sequences, in which the same outcomes can occur with variable frequencies:

(O) *Probabilistic causation as a function of single-case propensities*



where, even though the same "effects", $Y,\ -Y,\ \ldots$, are brought about by the same "causes", X, with constant propensity N from trial to trial, nevertheless, the limiting frequencies $m^1/n,\ m^2/n,\ \ldots$, with which these outcomes in turn would occur or would have occurred, if sequences of X-trials were or had been infinite, are not at all invariable, where their expectations over finite and infinite sequences of trials can be calculated on the basis of classic limit theorems available for statistical inference, including the Borel theorem, the Bernoulli theorem and the central limit theorem (Fetzer 1981, Ch. 5 and 9).

Because these "causes" are single-case properties and single-case trials of individual chance set-ups, they exemplify proximate mechanistic causation, rather than any species of teleological causation (intentional or otherwise), thereby establishing an appropriate foundation for an acceptable analysis of probabilistic causation as the desired account of "indeterministic causation". Unlike frequency- or long-run propensity-based accounts, this interpretation can explain how it is possible for "different effects" to be brought about by "the same cause", i.e., under the very same conditions, within a purely mechanistic framework. Indeed, this approach also fulfills the additional probabilistic desideratum which Skyrms (1980) and Eells (1983) have endorsed that, insofar as there is "no end" to the variation in singular outcomes that might be displayed during any particular sequence of singular trials, there should also be "no end" to the variations in limiting frequencies displayed by those sequences themselves. That this result follows from the single-case propensity conception, moreover, should be viewed as very reassuring: if it were adopted as a "condition of adequacy" for an acceptable account of indeterministic causation, this requirement alone would enforce a sufficient condition for excluding both frequency-based and long-run propensity-based conceptions.

These reflections also bear upon the methodological desideratum that an adequate interpretation of single-case propensities ought to analyse physical probabilities by means of concepts that are understood independently of quantitative explications, which Suppes introduced and Eells (1983) has pursued. Although each of the four accounts we have considered would fulfill this condition in a different fashion, all of them seem to depend upon a prior grasp of the independent conception of relative frequencies in finite sequences (or short runs) of trials. The frequency theories add the notions of limiting frequencies and of hypothetical sequences, while the propensity theories add the notions of long-run and of single-case dispositions. The conception of a disposition, like that of a limit, however, was with us already; so, unless this requirement improperly implies that non-extensional notions must be given extensional definitions (a contradictory imposition), the concept of a single-case probabilistic disposition would appear to be (at least) as acceptable as that of a hypothetical long-run limiting frequency, relative to this condition.

In his recent discussion of Salmon's shift from frequency-based to propensity-based conceptions, Humphreys (1986) has objected to the idea of introducing the single-case notion of a probability distribution "which is impossible to construe in actualist terms". His position here appears to be related to the thesis that Armstrong (1983) refers to as "actualism", namely: the view that properties and relations exist only if they are instantiated. However, if this understanding of his position is correct, then it ought to be rejected: were a steel ball rolled across a smooth surface and allowed to come to rest only finite times during its existence, it would be silly to suppose it had no propensities (or causal tendencies) to come to rest upon any other of its nondenumerable surface points – and similarly for other sorts of phenomena. We typically assign truth-values to assertions concerning what would be or what might have been in probabilistic as well as non-probabilistic contexts with respect to actual and merely possible arrangements and contrivances. If all of these "effects" had to be displayed in order to "exist" (to be "real"), the world would be a vastly different – less threatening and promising – place.

Nonetheless, careful readers may discern a flaw in some of these examples, insofar as games of chance (such as tosses of coins) and other set-ups (such as rolls of balls) may be inadequate illustrations of indeterministic cases. While probabilistic causation appears to operate at the macro- as well as at the micro-level of phenomena, discriminating genuinely indeterministic cases from deterministic cases requires calibration for every property whose presence or absence makes a difference to the occurrence under consideration, a stringent epistemic condition that may remain unsatisfied for specific cases. Indeed, while our diagrams have been restricted to hypothetical extensions of the world's actual history in conveying the differences between these alternative conceptions, an adequate account of "indeterministic causation" carries with it the consequence that the history of an indeterministic world might be indistinguishable from the history of a deterministic world, insofar as they might both display exactly the same relative frequencies and constant conjunctions – where their differences were concealed "by chance" (Cf. Fetzer 1983).

Indeterministic Causation: Concluding Reflections

Perhaps it should be clear by now why the comprehensive reconciliation of mechanistic explanations with indeterministic theories has not gone smoothly – perhaps even why the very idea of indeterministic causation has posed such an enormous conceptual problem. Indeed, the difference between the long-run and the single-case propensity conceptions is subtle enough to be invisible to the unaided eye, not only because neither diagram (M) nor diagram (K) incorporates its (implicit) distant "causal arrows", respectively, but also because diagram (M), like diagram (O), only exhibits (explicit) proximate mechanistic "causal arrows". But this is as it should be, for matters are more subtle than that: proximate teleological$_2$ causation, after all, is a species of proximate mechanistic causation! Still, this seems reason enough to sympathize with those who have experienced great difficulty in locating these differences (such as Levi 1977, 1979). Thus, by way of closing, it might be worthwhile to invite attention to three of the most important lessons that we have here to learn.

The first concerns the relationship between mathematics and the physical world; for, although frequency- and propensity-based conceptions of causation satisfy the "probabilistic conditions" of admissibility, of ascertainability, and of applicability, they do so in strikingly different ways. Frequency constructions formalize causal relations as a variety of conditional probability, while propensity conceptions formalize them as a species of causal conditional instead. The difficulties generated by symmetrical formulations reinforce the necessity to preserve the distinction between pure and applied mathematics; for if causation could be "probabilistic" only by satisfying inverse as well as direct probability relations, there could be no adequate conception of "probabilistic causation" (whether it were frequency-based or propensity-based in kind).

The second concerns the relationship between causation and the history of the physical world. What we have sought and what we have found is a conception of probabilistic causation that can be applied no matter whether the history of the physical world happens to be short or happens to be long. There appear to be inherent advantages in analysing "probabilistic causation" as a single-case conception that might be directly applied to finite and to infinite sequences, rather than as a long-run conception that might be applied indirectly to singular and to finite sequences instead. For it would be foolish to suppose that causal relations can obtain only if the world's history is infinitely long or that singular events can "cause" singular "effects" only if specific limiting frequencies happen to obtain during (actual or hypothetical) world histories.

The third and final (but by no means least important) concerns the relationship between indeterministic and deterministic theories and laws; for the considerations that have been brought to bear upon the analysis of indeterministic causation also apply to the analysis of deterministic causation. Deterministic causation, after all, is no more "symmetrical" than is indeterministic causation; and deterministic "causes" should no more be defined in terms of infinite classes of singular events than should indeterministic "causes": both should apply no matter whether the world's history is short or is long! The demise of statistical causality and the rise of probabilistic causation, therefore, should reap added dividends in analysing deterministic theories as well, further enhancing our understanding of the causal structure of the world.

Acknowledgments I am grateful to Paul Humphreys and especially to Ellery Eells for criticism. I have also benefitted from the stimulating comments of an anonymous referee.

References

Armstrong DM (1983) What is a law of nature? Cambridge University Press, Cambridge

Eells E (1983) Objective probability theory. Synthese 57:387–442

Fetzer JH (1971) Dispositional probabilities. In: Buck R, Cohen R (eds) PSA 1970. D Reidel, Dordrecht, pp 473–482

Fetzer JH (1974) Statistical probabilities: single case propensities vs long run frequencies. In: Leinfellner W, Kohler E (eds) Developments in the methodology of social science. D Reidel, Dordrecht, pp 387–397

Fetzer JH (1981) Scientific knowledge. D Reidel, Dordrecht

Fetzer JH (1983) Transcendent laws and empirical procedures. In: Rescher N (ed) The limits of lawfulness. University Press of America, Lanham, pp 25–32
Fetzer JH (1987) Critical notice: Wesley Salmon's scientific explanation and the causal structure of the world. Philos Sci 54:597–610
Good IJ (1961/62) A causal calculus I–II. Brit J Philos Sci 11:305–318 and 12:43–51
Hempel CG (1962) Deductive-nomological vs statistical explanation. In: Feigl H, Maxwell G (eds) Minnesota studies in the philosophy of science. University of Minnesota Press, Minneapolis, pp 98–169
Hempel CG (1965) Aspects of scientific explanation. The Free Press, New York
Hempel CG (1968) Maximal specificity and lawlikeness in probabilistic explanation. Philos Sci 35:116–133
Humphreys P (1986) Review of Wesley Salmon's scientific explanation and the causal structure of the world. Found Phys 16:1211–1216
Levi I (1977) Subjunctives, dispositions, and chances. Synthese 34:423–455
Levi I (1979) Inductive appraisal. In: Asquith P, Kyburg H (eds) Current research in philosophy of science. Philosophy of Science Association, East Lansing, pp 339–351
Salmon WC (1967) The of scientific inference. University of Pittsburgh Press, Pittsburgh
Salmon WC (1970) Statistical explanations. In: Colodny R (ed) The nature and function of scientific theories. University of Pittsburgh Press, Pittsburgh, pp 173–231
Salmon WC (ed) (1971) Statistical explanation and statistical relevance. University of Pittsburgh Press, Pittsburgh
Salmon WC (1975) Theoretical explanations. In: Korner S (ed) Explanation. Basil Blackwell, Oxford, pp 118–145
Salmon WC (1978) Why Ask, 'Why?' Proc Addresses Amer Philos Assoc 51: 683–705
Salmon WC (1980) Probabilistic causality. Pac Philos Quart 61:50–74
Salmon WC (1984) Scientific explanation and the causal structure of the world. Princeton University Press, Princeton
Skyrms B (1980) Causal necessity. Yale University Press, New Haven
Suppes P (1970) A probabilistic theory of causality. North-Holland, Amsterdam
van Fraassen B (1980) The scientific image. Oxford University Press, Oxford
von Mises R (1964) Mathematical theory of probability and statistics. Geiringer H (ed) Academic, New York

Part II
The Objectivity of Macro-Probabilities

Chance and Necessity:
From Humean Supervenience to Humean Projection

Wolfgang Spohn

Introduction

Probability abounds in the natural and social sciences. Yet, science strives for objectivity. Scientists are not pleased when told that probability is just opinion and there is no more sense to it. They are prone to believe in objective probabilities or chances. This is an essay about how to understand them.

Indeed, it is my first serious attempt in English[1] to come to terms with the notion of chance or objective probability. I cannot help feeling that this is a presumptuous enterprise. Many great minds have penetrated the topic. Each feasible position has been ably defended. No philosophically relevant theorem remains to be discovered. What else should there be to say? Yet, the issue is not settled. Even though all pieces are on the table, no one missing, how to compose the jigsaw puzzle is still not entirely clear. Philosophical uneasiness continues. Everybody has to try anew to put the puzzle together. So, here is my attempt to do so.

Let me lay my cards on the table right away. An event, or a state of affairs, is *chancy* iff it is partially determined by its past, to some specific degree; some might call this an Aristotelian conception of chance. Chance laws, then, generalize over such singular partial determinations. Likewise, a state of affairs is *necessary* (in the sense not of metaphysical, but of *natural* necessity) iff it is fully determined (i.e., sufficiently caused) by its past.[2] Deterministic laws generalize over such singular full determinations, so that we may reversely say that a state of affairs is necessary iff it is entailed by the laws and its past. This parallel will become important later on.

There *is* determination. There are deterministic laws, or so we believed at least for ages. And there are chance laws and hence chancy events, as modern physics tells us. Objective probabilities may thus be conceived as single-case propensities of a radical kind: propensities of the entire world as it has developed up to now to realize not only this or that current state of affairs, but in effect this or that entire future evolution.[3] The localization of propensities is a secondary, though, of course, important issue. The primary and really vexed issue is how at all to understand partial and full determination.

W. Spohn (✉)
Department of Philosophy, University of Konstanz, 78457 Konstanz, Germany
e-mail: wolfgang.spohn@uni-konstanz.de

E. Eells and J.H. Fetzer (eds.), *The Place of Probability in Science*, Boston Studies in the Philosophy of Science 284, DOI 10.1007/978-90-481-3615-5_5,

Given that there is partial determination, subjectivism or eliminativism concerning objective probabilities, a position associated with Bruno de Finetti and his positivistic predilections, is out of place. (Still, the most basic truths lie in his insights, and this essay will end up as hardly more than a projectivistic reinterpretation of de Finetti's views.)

Reductionism concerning objective probabilities seems ill-guided, too, whether in the analytical form trying to define chances in non-probabilistic terms as, e.g. (hypothetical) frequentism does or in the weaker ontological form as displayed in the doctrine of Humean Supervenience championed by David Lewis. Indeed, the failure of Humean Supervenience is nowhere clearer, I find, than in the case of chances.

Hence, realism without reductionism is perhaps what we should be heading for. I am indeed attracted by the picture as sketched, e.g., by Black (1998, pp. 371f.) who argues against Lewis that the world is more than "a vast mosaic of local matters of particular fact" (Lewis 1986, p. ix), more, as it were, than a pattern of inert colors; it is also a pattern of pushes, hard deterministic as well as soft chancy ones. Maybe we should accept realism about primitive laws, dispositions, capacities, propensities, etc. (or their categorical bases), as has been vigorously defended by Armstrong (1983, in particular Chapter 9, and 1997, Chapter 15) and in quite a different way by Cartwright (1989).

Yet I share the widespread epistemological concerns about Australian realism that are as old as Hume's criticism of necessary connexion or determination. What we need to get explained, at least, is the theoretical web within which chances get their role to play.[4] However, the explanations given by propensity theorists are generally not in good shape.[5] And so I appear to be torn by my various dissatisfactions, finding no place to rest.

No other than David Hume has suggested a position possibly comforting everyone, with his doctrine that causation is an idea of reflexion and that necessary connexion is nothing but determination or customary transition *in thought*. The doctrine has received its most extraordinary shape in Kant's transcendental idealism. Nowadays, it is rather called projectivism and defended by Simon Blackburn under the label 'quasi-realism' and summarized thus:

> Suppose we honor the first great projectivist by calling 'Humean Projection' the mechanism whereby what starts life as a non-descriptive psychological state ends up expressed, thought about, and considered in propositional form. Then there is not only the interest of knowing how far Humean Projection gets us. There is also a problem generated even if the mechanism gets us everywhere we could want. If truth, knowledge, and the rest are a proper upshot of Humean Projection, where is it legitimate to invoke that mechanism? Perhaps everywhere, drawing us to idealism, or nowhere, or just somewhere, such as the theory of value or modality. (1993, p. 5)

This 'mechanism', I shall argue, is operative at least in the case of chance and natural necessity. It is thus no accident that I am referring twice to Hume within one page. The move from Humean Supervenience to Humean Projection will be *our* move in this paper. (Indeed, I find that the latter is much better anchored in Hume's writings than the former.)

The crux of projectivism, though, is that it may sound attractive as a general strategy, while remaining poor in constructive detail. Thus it is not likely to satisfy the probability community. Indeed, if one looks at recent surveys such as Gillies (2000), projectivism does not figure there under its own or any other name. This is the point where I hope to add a bit to the present discussion.[6]

As the reader may have guessed, this paper is largely an argument with David Lewis' philosophy of probability. This has a personal motive. I well recall how enthralled I was by Lewis (1980) – and how bewildered by the continuation in Lewis (1986, Introduction and Postscripts to 1980, and 1994). I just had to come to grips with his work. There is also a substantial reason. Lewis' account is peculiarly ambiguous. He starts inquiring the epistemology of chance and ends up investigating its ontological grounds. Thus, I find it most instructive to follow his line of thought and to search for the point of departure for a more adequate account.

There is a third reason. The parallel between deterministic and chance laws is obvious; it would be awkward to account for them in a wholly disparate manner. Lewis expressly pursues this parallel; after apparent success in the deterministic case, his strategy just had to carry over to chance laws, as elaborated in his (1994). Therefore, Lewis is the natural point of contact on this score, too, and however I diverge from Lewis' account of chance, the divergence must work for deterministic laws as well. In fact, I see how it will. Contrary to appearances, natural necessity or full determination or lawlikeness is still less understood than partial determination; even the appropriate analytical means were missing. The theory of ranking functions will bring progress here. This remark, though, will be briefly outlined, and can be more easily grasped, I hope, after treating the actually more familiar probabilistic case.[7]

The paper will proceed in the following way: We shall start in the section *Chance-Credence Principles* with recapitulating the central role the Principal Principle has, according to Lewis, for understanding chance. Lewis gives substance to this principle by claiming admissibility, as he calls it, for historical and chance information; this will be discussed and simplified in the section *The Admissibility of Historic and Chance Information*. The admissibility of chance information drives him into a contradiction, though, with the doctrine of Humean Supervenience. Lewis proposes to reform the Principal Principle, but I shall argue in the section *The Admissibility of Chance Information and Humean Supervenience* that it is rather Humean Supervenience that has to go. This provokes a closer look at that doctrine, and we shall see in the section *Humean Supervenience* that it is inherently questionable. So, this will be the point where a projectivistic reconstruction of the notion of partial determination is likely to deliver a more coherent account. The reconstruction will be carried out in the section *Projection Turns the Principal Principle into a Special Case of the Reflection Principle*, via the observation that the Principal Principle may be taken, in a precise way, as a special case of the Reflection principle propagated by van Fraassen (1984); this is no deep formal insight, but of some conceptual interest. The section *Humean Projection* will sum up the projectivistic doctrine and argue that it can meet familiar objections and serve the purposes for which Lewis had invoked Humean Supervenience. As explained, the whole line of reasoning must somehow carry over from chance to natural necessity or from partial to full determination. The appendix will indicate how this might go.

Chance-Credence Principles

Let us approach our topic, objective probability, via the Principal Principle, which seems to its baptizer "*to capture all we know about chance*" (Lewis 1980, p. 266, my emphasis) – a proper starting point, if this claim were true. There is in fact not only one principle relating chance and credence; subsequent literature has discerned a whole family of principles, which we do well to survey. So, let us start in a purely descriptive mood; we shall become involved into debate soon enough.

The basic idea relating chance and credence is very old and familiar; it is simply that if I know nothing about some proposition A but its chance, then my credence in A should equal this chance. This is the *Minimal Principle* (as Vranas 2004 calls it):

$$\text{(MP)} \quad C(A|P(A) = x) = x.$$

Here, C stands for subjective probability or credence (the association with Carnap's 'confirmation' is certainly appropriate), and P for objective probability or chance (or propensity, if you like). The subject having the credence remains unspecified, since (MP) is, as it were, a generic imperative; (MP), like the subsequent principles, is a rationality postulate telling us how any credence function should reasonably behave.

(MP) is the starting point of the sophisticated considerations in Lewis (1980). (MP) is also called "Miller's Principle", because Miller (1966) had launched a surprising early attack on it. However, (MP) is not an invention of the recent philosophical debate. It is known for long also under the label "direct inference".[8] In fact, it is implicit in each application of Bayes' theorem to statistical hypotheses; there the 'inverse' posterior probabilities of the hypotheses given some evidence are calculated on the basis of their prior (subjective) probabilities and the 'direct' probabilities or likelihoods of the evidence under the hypotheses; and these 'direct' probabilities hide an implicit use of (MP). The merits of the recent discussion pushed by Lewis (1980) and others are rather to scrutinize variants of (MP).

Before proceeding to them there are, however, various things to clarify. Philosophy first, I propose. If Lewis is right that principles like (MP) "capture all we know about chance", then the philosophical interest of these principles is evident. Lewis does not really argue for this claim. In fact, he does not make it, it only seems true to him. Indeed, he cannot strictly believe it by himself. When, as we shall see later on, he claims chances to Humeanly supervene on particular facts, then he clearly transcends the Principal Principle. And I shall end up agreeing with Arntzenius, Hall (2003) that there must be more we know about chance.

The point should rather be seen as a challenge. For, what is true is Lewis' assertion "that the Principal Principle is indeed informative, being rich in consequences that are central to our ordinary ways of thinking about chance" (1980, p. 288), as is amply proved in his paper. For instance, it follows that chance conforms to the mathematical axioms of probability. The challenge then is what else there might be to say about chance. In default of an explicit definition of chance we seek for an implicit characterization, and it seems that we have already gone most of our way with the extremely neat Minimal Principle (which, as we shall see, is hardly strengthened by the other principles still to come). The more we are captivated by the Principal Principle, the harder the challenge.

The harder, though, also the philosophical puzzle posed by chances. It is strange that chances that are supposed to somehow reflect objective features of the external world should be basically related to our beliefs in a normative way. Our implicit characterizations of other theoretical magnitudes do not look that way. And the more weight is given to this relation, the more puzzling it is. How should we understand the peculiar normative power of that objective magnitude to direct our beliefs? If, indeed, the Principal Principle is all we know about chance, that power turns into sheer miracle. Why should we be guided by something the only known function of which is to guide us? Preachers may tell such things, but not philosophers.[9] One of Lewis' motives for the doctrine of Humean Supervenience is, we shall see, to solve this puzzle; indeed, he claims it to be the only solution. We need not take a stance right now, but we must always be aware in the subsequent discussion of the basic merits and problems of the Principal Principle. We are dealing here with high philosophical stakes.

At the moment, though, we must be a bit more precise about (MP). First, we must be clear about the domains of the functions involved. The chance measure P takes propositions, I said. We should not start, though, philosophizing about propositions. Let us simply assume that propositions are subsets of a given universal set W of possible worlds.

Is any kind of proposition in the domain of P? This is an open question. It is debatable which propositions may be chancy or partially or fully determined and which not. There may be entirely undetermined propositions and there may be propositions for which the issue makes no sense. Let us leave the matter undecided and grant, in a liberal mood, that at least all *matters of particular fact*, and hence all propositions algebraically generated by these facts, have some degree of determinateness, i.e., chance. Lewis (1994, pp. 474f.) has an elaborate view on what particular facts are; here we may be content with an intuitive understanding.

In any case, a proposition saying that some factual proposition has some chance is not a particular fact in turn. This does not exclude that such a chance proposition is algebraically generated by particular facts, but neither does it entail it; it is crucial for this paper not to presuppose from the start that chance propositions are factual in the same way as particular facts. So, let us more specifically assume that each $w \in W$ is a complete possible course of particular facts. Whether we should be more liberal concerning the domain of chance will be an issue later on.

Credence is not only about particular facts, but also about possible chances; this is explicit in (MP) itself. Thus, if $\mathcal{P}$ denotes the set of possible chance measures for W, then $W \times \mathcal{P}$ is the possibility space over which the credence C spreads.

Moreover, I shall be silent about the precise algebraic structure of the set of propositions[10] and just assume that each $P \in \mathcal{P}$ is defined on some algebra over W and C on some algebra over $W \times \mathcal{P}$. Accordingly, I shall be silent about the measures we are considering being finitely or σ-additive. This sloppiness will have costs, but rigorous formalization would have costs, too. I am just following the practice usually found in the literature I am mostly referring to.

For instance, one consequence of sloppiness is that (MP) does not make strict sense, since the condition will usually have probability 0. Lewis says that we should move then to non-standard analysis and hyperfinite probability theory where the condition in (MP) may be assumed to have infinitesimal probability. More easily said than done. Within standard probability theory one may circumvent the problem by stating (MP) in the more general form:

$$\text{(MPI)} \quad C(A|P(A) \in I) \in I \text{ for any open interval } I.$$ [11]

This issue will return, and all the principles I am going to discuss should be restated accordingly.

There is another reason why (MP) will not do as it stands. C may not be any credence function. If C is already well informed about A, for instance by being based on the observation of A or of some effects of A, (MP) is obviously inadequate. As Lewis (1980, pp. 267f.) explains, this concern is excluded for sure only if C is an initial or a priori credence function, as conceived as the target of further rationality constraints also by Carnap in his inductive logic. To indicate this, I shall denote an a priori credence by C_0 (0 being a fictitious first time).

Finally, in order to present Lewis' ideas we must note that chance evolves in time; this is particularly clear when chance is conceived as partial determination. Even full determination evolves in time, unless determinism holds and everything is fully determined at all times. Moreover, chance is world dependent; how chance evolves in time may vary from world to world. In order to make these dependences explicit we must replace P by P_{wt}, the chance *in w at t*. Thus we arrive at a slightly more explicit version of the Minimal Principle:

$$\text{(MP*)} \quad C_0(A|P_{wt}(A) = x) = x.$$ [12]

Having said all this, let us return to our descriptive path through the family of chance-credence principles (cf. also the useful overview in Vranas 2004). A first minor step proceeds to a conditional version of (MP) introduced by van Fraassen (1980, pp. 106f.), the *Conditional Principle*:

$$\text{(CP)} \quad C_0(A|B \;\&\; P_{wt}(A|B) = x) = x,$$

saying that, if you know nothing about A but its chance conditional on B, your conditional credence in A given B should equal this chance. (CP) is certainly as evident as (MP). We shall soon see that (CP) is hardly stronger than (MP).

David Lewis has taken a different, apparently bigger step. After retreating to the a priori credence C_0 in (MP) that is guaranteed to contain no information overriding the conditional chance information, Lewis poses the natural question which information may be added without disturbing the chance-credence relation stated in (MP). He calls such additional information admissible, and thus arrives at what he calls the *Principal Principle*:

$$\text{(PP)} \quad C_0(A|E \;\&\; P_{wt}(A) = x) = x \text{ for each admissible proposition } E.$$

But what precisely is admissible information? The answer is surprisingly uncertain; the literature (cf. e.g., Strevens 1995 and Vranas 2004) strangely vacillates between defining admissibility and making claims about it. I think it is best to start with a clear definition, which is obvious, often intimated (e.g., by Vranas 2004, Footnote 5), but rarely endorsed in the literature (e.g., by Rosenthal 2004, p. 174):

(DefAd) E is *admissible wrt A given D* iff $C_0(A|E \cap D) = C_0(A|D)$. Specifically, E is *admissible wrt A in w at t* iff E is *admissible wrt A* given $P_{wt}(A) = x$.

The first general part says that E is admissible wrt A given D iff E does not tell anything about A going beyond D according to the a priori credence C_0. Admissibility is nothing but *conditional independence*. The second part gives the specification intended and used in (PP).

Obviously, the definiens states at least a necessary condition of admissibility; any admissible E not satisfying this condition would directly falsify (PP). I propose to consider the necessary condition to be sufficient as well. This strategy trivializes (PP); with (DefAd), (PP) reduces to nothing more than (MP), and the issue of admissibility is simply deferred. Still, I find the detour via (DefAd) helpful. It clearly separates the meaning of admissibility from the substantial issue which propositions E should be taken to satisfy (DefAd). This issue is our task in the next section.

One may still wonder why one should take the necessary condition for admissibility to be also sufficient. We may have stronger intuitions concerning admissibility. We may, for instance, think that two pieces of information admissible individually should also be jointly admissible, a feature not deducible from (DefAd). Or we may think that any E admissible wrt A in w at t should be so for general reasons pertaining to w and t and not to idiosyncratic reasons pertaining to A. And so on. However, the theoretical tactics is always to start with the weakest notion, which is (DefAd) in the case at hand. The substantial claims about admissibility will then also take a weak form, but we are always free to strengthen them. The point is that we would not have the reverse freedom when starting with a stronger notion right away.[13] A further worry may be that (DefAd) lets a priori credence C_0 decide about admissibility. However, we should read (DefAd) the other way around; whatever the substantial claims about admissibility, they are constraints on C_0 via (DefAd).

The Admissibility of Historic and Chance Information

Lewis (1980) makes two substantial claims about admissibility. The first is that each piece of historic information is admissible. If I know the chance $P_{wt}(A)$ that A has in w at t, this knowledge cannot be improved by any information about what happened in w up to t; $P_{wt}(A)$ summarizes, as it were, all there is to know in w up to t. Let us denote the history of the world w up to time t by H_{wt}. $H_{wt} = \{v \in W | H_{vt} = H_{wt}\}$ is a proposition. Moreover, let us say that the proposition E is only abouthistory up to t, or *t-historical*, for short, iff for each world w either $H_{wt} \cap E = \varnothing$ or $H_{wt} \subseteq E$. Then Lewis' claim is:

$$\text{(AdH)} \quad \textit{If } E \text{ is } t\text{-historical, then } E \text{ is admissible wrt } A \text{ in } w \text{ at } t.$$

Note that reference to A is empty; only the relation of E to t is relevant to (AdH).

This claim is almost universally accepted. Lewis (1980, p. 274) himself raises a doubt about (AdH). Could there not be a crystal ball that foretells me for sure whether or not A happens, even if A is chancy? I shall explain later why I am not worried by this alleged possibility. Here, I just join (AdH). Thus, the Principal Principle starts unfolding some strength. Let me add three remarks that deepen the understanding of the point.

First, Lewis (1980) presents the case as if the admissibility of historical information were specifying (PP) and thus rationally constraining credence. So it does, but the core of the matter is thereby obscured in my view. (PP) and (AdH) immediately entail what might be called the *Determination Principle*:

$$\text{(DP)} \quad P_{wt}(H_{wt}) = 1.$$

(This follows by replacing A as well as E in (PP) by H_{wt}.) (DP) simply says that history is no longer chancy. This consequence is, of course, intended. However, it is not about credence, but only about chance. In fact, it is an analytic truth about (partial) determination: what is past is fully determined.[14] Hence, it is more illuminating to realize that only this analytic truth needs to be added to (CP) to entail (AdH).[15] The matter will further simplify later in this section.

The second point is one I have not seen emphasized in the Principal Principle literature, though it deems important to me: In the prolonged efforts of understanding objective probabilities, whether as frequencies or propensities, the so-called *reference class problem* stood out as central and embarrassing.[16] The probability of a particular event seemed to depend on the reference class within which it was considered. Thus, that event could be assigned an objective probability only if one could distinguish the objectively correct reference class, apparently a dubious matter. The general recommendation was to rely on the narrowest reference class (or on the broadest reference class equivalent to the narrowest one); see also Hempel's so-called criterion of maximal specificity. This may be taken as the narrowest *available* reference class; but availability imports epistemic relativity. Or one may engage into the difficult business of Salmon (1984, pp. 60ff.) of distinguishing *objectively* homogeneous reference classes. If Lewis is to explain us objective probabilities, he must respond to this problem.

He does implicitly, his response is (AdH). For the objective probabilities at t the whole history, H_{wt}, *is* the objectively narrowest reference class; there could not be any more specific one. Indeed, it is hardly a class; it has only one member, actually unrepeatable and only counterfactually repeatable. Hence, this is at best a trivial and purely conceptual solution of the reference class problem; it does not even touch the real and deep methodological problem to specify sound and manageable reference classes. However, this is a problem the philosopher must leave to the scientist; the philosopher can only say what the ultimate standard is with which to compare all actually considered reference classes.

The third remark is related. If all the history up to t is admissible wrt some proposition A in w at t, this means that up to t there is absolutely nothing more to know about A than its chance. You can learn absolutely everything up to t and you will be none the wiser concerning A. If you do not even know the chance of A, you are even more in the dark; the chance of A at t is the best you can know about A at t. This is the core intuition about partial determination: if A is in some way partially determined at t, there is nothing before t that would determine A in any other way. And knowledge before t can at best equal determination at t.

This point is reflected in many accounts of probability, for instance in the old idea that genuine random processes cannot be outfoxed by a gambling system. The same thought is found in von Mises' (1919) definition of a collective as a sequence for which no place selection results in a subsequence with a deviating limit of relative frequencies and in the subsequent explications of this approach with recursion and complexity theoretic means (cf. Church 1940; Chaitin 1966). Salmon (1984, pp. 60ff.) realizes the same basic idea in terms of his objectively homogeneous reference classes, though in a different way. It is important to see all this connected with (AdH).

Let us turn to the second kind of admissible information acknowledged by Lewis (1980): information about the chances themselves, not only about the actual ones, but also about the ones as they would have been at various times. If I know the actual chance of A, how could information how other chances would have been tell me more about A? It cannot, as Lewis (1980, pp. 274ff.) argues.

To state this more precisely: Let T_w be the complete theory of chance holding at the world w, i.e., according to Lewis, the conjunction of all conditionals true at w having the form: "if the history of w up to t' had been $H_{vt'}$, then $P_{vt'}$ had been the chances in w at t'." Lewis assumes T_w to be a proposition over W; this is a controversial assumption to be discussed later. Is it at all a proposition over $W \times \mathcal{P}$ (or $W \times \mathcal{P}^T$ – cf. footnote 12)? Prima facie not, since the counterfactual conditional is not among the Boolean operations. Still, we may take T_w to be in the domain of C_0; the issue will be cleared up on the next page. Furthermore define E to be a *chance proposition* iff for each world w either $T_w \cap E = \varnothing$ or $T_w \subseteq E$. Then Lewis' second admissibility postulate is:

(AdP) If E is a chance proposition, then E is admissible wrt A in w at t.

Note again that the reference to A and even to t is empty; all that matters is that E is a chance proposition.

Lewis assumed that separate admissibility of historic and chance information entails their joint admissibility. This does not follow, however, on the basis of (DefAd). Hence, we should better read (AdH) or (AdP) as saying for each that its kind of information is admissible *given* the other kind of information; this entails its unconditional admissibility.[17]

At this point, we can easily see that the Conditional Principle (CP) is hardly stronger than the Minimal Principle (MP). If $P_{wt}(A \cap B) = y$ is admissible wrt B in w at t and $P_{wt}(B) = z$ is admissible wrt $A \cap B$ in w at t, then (PP) yields $C_0(A \cap B | P_{wt}(A \cap B) = y \;\&\; P_{wt}(B) = z) = y$ and $C_0(B | P_{wt}(A \cap B) = y \;\&\; P_{wt}(B) = z) = z$ and both together yield (CP). In other words: We have to add to (MP) only the admissibility of a tiny bit of chance information in order to get (CP).

(PP) + (AdH) + (AdP) may finally be combined to what Lewis (1980) called the Principal Principle reformulated and was later called the *Old Principle*, since it is not yet the end of the story.

$$\text{(OP)} \quad C_0(A | H_{wt} \;\&\; T_w) = P_{wt}(A).$$

This follows from (PP) because H_{wt} & T_w is admissible according to (AdH) and (AdP), T_w contains "if H_{wt}, then $P_{wt}(A)$ is the chance of A in w at t", and thus H_{wt} and T_w entail what $P_{wt}(A)$ is. Conversely, (OP) entails (PP) + (AdH) + (AdP). So, (OP) is a very elegant summary of the foregoing discussion.

The story can be further simplified, though. Let us look at T_w again. It is not quite clear why it has to take the specific complicated form, perhaps because T_w is to allow that some histories leave some events not even partially determined. However, we wanted to ignore such complications and assumed that all matters of particular fact are partially determined (or almost fully determined via chance 1). Hence, T_w claims for each possible history H_{vt} a full chance measure P_{vt} for W. Then, however, we may condense the whole theory T_w into one big chance measure P_w such that the time-dependent chance $P_{w,vt}$ derives from P_w through conditioning by H_{vt}.[18] We thus simply replace the conditionals with probabilistic consequents by conditional probabilities.[19] That it is possible to so condense T_w is indeed a consistency requirement for T_w, which becomes explicit also in Lewis (1980, pp. 280f.) in his discussion of the kinematics of chance.[20] P_w thus is the *time-independent chance law* or *scheme of partial determination* as it holds in w for all propositions over W, and T_w simply says that P_w is as it is.[21] Hence, we have arrived at the following reduction of Lewis' terminology:

$$\text{(RED)}\, T_w = \{P_w\} \text{ (or rather } = W \times \{P_w\} \subseteq W \times \mathcal{P}),$$
$$\text{and } P_{w,vt}(A) = P_w(A | H_{vt}) \text{ for all } A,\ v, \text{ and } t.$$

This shows at the same time that T_w is indeed a proposition over $W \times \mathcal{P}$ (indeed over $\mathcal{P}$ alone) and is thus in the domain of C_0 as we have originally conceived it.

(RED) makes clear that all the considerations about time-dependent chance are perhaps intuitively helpful and perhaps required for more general chance theories, but merely a conceptual detour within our frame. (RED) also explains why the above Determination Principle is analytic; (DP) follows from the definition (RED). And (RED) reinforces the redundancy of (AdH); given (RED) and (AdP) (OP) is just an application of the Conditional Principle (CP). However, we just saw that (CP) is entailed by (MP*) and (a small part of) (AdP). So, the latter two are the only basic assumptions we need. (RED) finally helps us to express the Old Principle still more simply:

$$\text{(OP}^*) \quad C_0(.|T_w) = P_w.$$

Indeed, (OP*) looks like the Minimal Principle itself; the only difference is that (MP*) refers to the chance of a single proposition, whereas (OP*) refers to a whole chance measure. It is only from the restricted perspective of (MP*) that (OP*) appears to additionally assume the admissibility of chance information. Initially, I suppose, intuition would have been indifferent between (MP*) and (OP*).

The Admissibility of Chance Information and Humean Supervenience

So far, so good. We might be happy with (OP*) and start discussing its philosophical significance. Alas, the story takes a most unexpected turn, for which it is important that we have discerned (AdP) as an additional assumption in (OP). (OP) thus becomes the starting point of considerable confusion. The source of the trouble is that Lewis not only takes chance-credence principles like (MP*) to provide the most basic understanding of chance, but also maintains the ontological doctrine of so-called Humean Supervenience – because this is an attractive metaphysical doctrine, and because such chance-credence principles seem to require it. The trouble is real, and therefore we shall have to scrutinize both grounds of Humean Supervenience. But let us first have a formal look at what the trouble is.

With respect to chance, Humean Supervenience consists in the claim:

(HS) T_w supervenes on the totality of particular facts in w.

With our reduction (RED) of T_w and our understanding of the worlds in W as mere totalities of particular facts, we might as well express this claim thus:

(HS*) P_w supervenes on w.

It is not quite clear for which worlds w (HS) is to hold. Certainly for the actual world we live in. One may think that (HS) applies to all worlds and is thus a necessary truth. Lewis (1994) sees it only as a contingent truth; (HS) is to hold only for worlds *like ours*, certainly a more modest and a more mysterious view. We do not have to take a stance here.

Since we are a bit sloppy concerning the algebra of propositions, we may say that (HS) amounts to the claim that T_w is identical to a proposition over W.[22] (HS) thus says there are not two possibility spaces, one for possible facts (forming the domain of chances) and one for possible chances (jointly forming the domain of credences).

The latter rather reduces to the former; there is only the space of possible facts. Chance propositions are in effect factual propositions – and thus in the domain not only of credences, but of chance measures themselves.

Now, however, we are caught in paradox. Imagine that our world w, after having started with H_{wt}, continues with some possible future F_{wt}. F_{wt} should have at least a tiny chance of coming about; so $P_{wt}(F_{wt}) > 0$ and, according to (OP), $C_0(F_{wt}|H_{wt} \,\&\, T_w) > 0$. On the other hand, F_{wt} may be an undermining future in the sense that $H_{wt} \cap F_{wt}$ ($= \{w\}$) is not in the supervenience base of T_w, i.e., according to H_{wt} and F_{wt} w would be governed by some chance law different from T_w. Then F_{wt} is impossible given H_{wt} & T_w, i.e., $C_0(F_{wt}|H_{wt} \,\&\, T_w) = 0$. To put the case very briefly with reference to (OP*): Consider the factual proposition $\overline{T}_w$ that T_w is false. Clearly, $P_w(\overline{T}_w) = C_0(\overline{T}_w|T_w) = 0$, However, if w is genuinely chancy, we should have $P_w(\overline{T}_w) > 0$. Somewhere, we have made a mistake.

It seems clear where. Given (HS), not all chance information can be admissible, since information about the future may well be inadmissible and since chance information *is* information about the future according to (HS).[23] Indeed, we should conclude that most chance information is inadmissible, though it is hard to be more precise because it is not so clear how supervenience exactly works, in which complex of particular facts T_w exactly consists.

However, as Lewis (1994) argues, most chance information is at least nearly admissible, and (PP) and (OP) work approximately well even under the assumption of (AdP); the mistakes we incur are below noticeability. Still, the question is: if (OP) is only approximately valid, what is the standard it approximates? Following Thau (1994) Lewis (1994) proposes that this standard is provided by the *New Principle*:

$$\text{(NP)} \quad C_0(A|H_{wt} \,\&\, T_w) = P_{wt}(A|T_w),$$

or in our reduced form:

$$\text{(NP*)} \quad C_0(.|T_w) = P_w(.|T_w).$$

This appears to solve our problem. The derivation of the paradox of undermining futures is blocked when we use (NP) instead of (OP), and the approximate validity of (OP) is explained by the fact that the difference between $P_{wt}(A)$ and $P_{wt}(A|T_w)$ is mostly below noticeability.

Is this an ad hoc solution? No. As Hall (1994, p. 511) and Strevens (1995, p. 557) observe and Hall (2004, pp. 104f.) insists, (NP) is a consequence of (CP) and (DP) which are uncontested.[24] Moreover, the admissibility of chance information that drove (OP) into paradox is guaranteed for (NP); T_w, and hence any weaker chance information, is trivially admissible wrt A given T_w & $P_{wt}(A|T_w) = x$. Hence, (NP) appears to be the right way to reconcile Humean Supervenience with (PP), the admissibility of historic information and the general *in*admissibility of chance information.

However, Lewis (1994) is not entirely satisfied. He says:

> A feature of Reality deserves the name of chance to the extent that it occupies the definitive role of chance; and occupying the role means obeying the old Principle, applied as if information about present chances, and the complete theory of chance, were perfectly admissible. Because of undermining, nothing perfectly occupies the role, so nothing perfectly deserves the name. But near enough is good enough. (p. 489)

And thus Lewis acquiesces in chances obeying (OP) not quite perfectly.

This remark provokes the final twist of the story. As Arntzenius, Hall (2003) point out, (NP) entails that there *is* a magnitude occupying the definitive role of chance perfectly, i.e. satisfying (OP) strictly. Suppose that the world w determines the chance theory T_w; according to (HS) it does so in some particular manner. And suppose that T_w allows for undermining futures so that (OP) does not apply. Now, define $P_w^* = P_w(.|T_w) \neq P_w$ and $T_w^* = \{P_w^*\}$. So, T_w and T_w^* obviously are incompatible chance theories – in one sense. However, change also the supervenience bases for chances; say for each w that it is not P_w, but rather P_w^* that is determined by (the facts of) w. So, in another sense, T_w is a factual proposition over W according to the initial way of determination, and T_w^* is so, too, according to the modified way of determination. And in this sense, they are not incompatible. On the contrary, T_w entails T_w^*, since whenever $T_v = T_w$ according to the initial way of determination, $T_v^* = T_w^*$ according to the modified one (though not necessarily vice versa). Moreover, T_w^* cannot be threatened by undermining futures. And, this is the upshot, if P_w, T_w satisfy (NP*), i.e., if $C_0(.|T_w) = P_w(.|T_w)$, then P_w^*, T_w^* satisfy (OP*), i.e., $C_0\left(.|T_w^*\right) = P_w^*$.[25]

Hence, if the old principle is definitive of the chance role, as Lewis says, then P_w^*, rather than P_w, should be the chance law governing the world w. If we tend to say P_w is determined by the particular facts, we should say it is rather P_w^* that is determined by those facts. Thus, we face a new paradox, at least if we think that true chance theories must allow for undermining futures. And even if we deny this and rather attempt to choose P_w right away so that $P_w^* = P_w$, then Arntzenius, Hall (2003) complete their argument by showing that chances then behave in an unacceptable way.

Schaffer (2003) tries to escape by claiming vagueness. Chance may be given by P_w or by P_w^*, and disambiguation is of little importance, since the difference is small, anyway. However, it is not chance that is vague, I think, only our thinking about it is not clear enough. My conclusion is that we are in deep trouble and have not found any stable position concerning the admissibility of chance information and the possibility of undermining futures. What got us there? It was, of course, the assumption of Humean Supervenience unquestioned so far. It is high time to attend to it more closely.

(HS) assumed that the chance proposition T_w over $\mathcal{P}$ is supervenient upon, or, with sloppy algebra, identical with some factual proposition over W. As a consequence, we had to consider chance propositions as being in the domain of P_w and P_{wt}, at least under a liberal, though not exceptional conception of this domain, and hence we had to consider such chances as $P_w(T_w)$ or whether or not $P_{wt}(A) = P_{wt}(A|T_w)$. By contrast, if we give up (HS), we are free to reject such expressions and in particular the New Principle as meaningless. If we do so, the admissibility of chance information is rescued from paradox and perfectly acceptable. Indeed, at so many places philosophy ran into trouble in the past decades with iterating (the same kind of) modality. We should have been warned.

I raised the point in my (1999, p. 170). But it is underrated in the literature. The worst Lewis (1994) and Hall (1994) say about (NP) is that it is messy and user-hostile. Arntzenius, Hall (2003, p. 175) only say that the non-reductionist rejecting (HS) is free to assume $P_{wt}(T_w) = 1$ and to thus eliminate the discrepancy between (OP) and (NP). Hall (2004, p. 99) insists that this stipulation is harmless. Formally, this is correct, but the non-reductionist need not even take this step. And he should not; the harm done consists in blurring the issue. It creates the impression that the issue between the reductionist and the non-reductionist would be whether $P_{wt}(T_w)$ is equal to or smaller than 1; it creates the delusion of there at all being a meaningful issue. It simply makes no sense to say that there is some chance that our world is governed by this scheme of partial determination rather than that or that this atom has (at t) a propensity of.4 of having (at t) a propensity of.2 of decaying (within the next hour).

Hoefer (1997, p. 328) expressly agrees by saying:

> The laws are what they are because of the pattern of events in history, and not what they are "by law". This is just a restatement of the core idea of Humean analyses of law. For just the same reason, the chances are not what they are "by chance", and the quantity $P_{tw}(T_w)$ should be regarded by a Humean as an amusing bit of nonsense.

However, his argument is a different one. He doubts that all particular facts (and their Boolean combinations) are in the domain of the chance function. Hence, even if the chance of chancy facts supervenes on particular facts, the supervenience base will usually not be in the domain of the chance function. NP would only be guaranteed to make sense for the Human supervenientist, if all particular facts were chancy. By contrast, I am granting the latter and arguing that NP still does not make sense.

Vranas (2004, p. 373) tries to save the "arguably dubious" assumption that chance propositions are in the domain of P_{wt}. He notices the potentially vicious circularity in such expressions as $P_{wt}(T_w)$, which is indeed a point of worry for the *non*-reductionist, but not for the reductionist, and he proposes to make sense of such expressions within reflexive situation theory (cf. Barwise and Etchemendy 1987) and thus ultimately within set theory without the foundation axiom. But why at all should the non-reductionist try to overcome his worry and take recourse to such remote means? For the non-reductionist particular facts and Boolean combinations thereof are chancy and what lies outside this domain is not. It is up to the reductionist to give an argument for conceiving the domain more broadly, and the argument must not presuppose (HS), as one would if one praises the apparent progress from (OP) to (NP).

Humean Supervenience

So, let us squarely face Humean Supervenience itself. I propose first to look at how Lewis thinks it is feasible. Once we shall have seen the doubtfulness of Lewis' construction I can proceed with an alternative account and then with a brief discussion of Lewis' reasons for taking (HS) to be without good alternative.

The thesis of Humean Supervenience says, according to Lewis (1994, p. 474)

> that in a world like ours, the fundamental relations are exactly the spatiotemporal relations ... and ... that in a world like ours, the fundamental properties are local qualities. ... Therefore it says that all else supervenes on the spatiotemporal arrangement of local qualities throughout all of history, past and present and future.

Because this holds only for worlds *like ours* (HS) is contingent. Should alien qualities in Lewis' sense play a role – irreducible chance would be such an alien quality –, the case may be different.

The bite of this claim emerges when we consider all the things that are extremely thorny for philosophers: laws, counterfactuals, causation – and objective probabilities. All this must be determined by the totality of particular facts, according to (HS). How? The crucial link is constituted by what Lewis calls the best-system analysis of law, which he takes over from F. P. Ramsey:

> Take all deductive systems whose theorems are true. Some are simpler, better systematized than others. Some are stronger, more informative, than others. These virtues compete: an uninformative system can be very simple, an unsystematized compendium of miscellaneous information can be very informative. The best system is one that strikes as good a balance as truth will allow between simplicity and strength. How good a balance that is will depend on how kind nature is. A regularity is a law iff it is a theorem of the best system. (Lewis 1994, p. 478)

So far this applies only to deterministic laws. But Lewis suggests to expand the best-system analysis to cover chance laws as well, and he makes clear that the inclusion of chance laws in the best system is primarily governed by relative frequency and symmetry. Some say that Lewis' position thereby basically reduces to frequentism, others say that it essentially transcends frequentism. We need not decide. We may well accept the best-system analysis for the time being. It is plausible, as far as it goes; it is, to echo Lewis, simple, but uninformative.

There are two critical points, though. The first is that the team of the best-system analysis and the Principal Principle introduces not only an ontological, but also an epistemological double standard. We have already seen the ontological double standard. The best-system analysis somehow establishes T_w as the chance theory true of w, whereas (OP) rather requires T_w^* to be determined by w. In addition, we now face an epistemological double standard. On the one hand, our beliefs aim at the best system guided by standards of simplicity and strength and their balance. On the other hand, one should think that all these standards are encoded in the a priori credence function C_0 that we seek to constrain by (OP) and other rationality postulates. I do not see an incoherence here, but neither do I see how the two standards go together or what results from the circular procedure of letting C_0 decide about the best system and feeding in the decision into condition (OP) on C_0. These are unresolved frictions, to say the least.[26]

The second critical point is, of course, whether the best-system analysis can at all bolster up Humean Supervenience about laws and chance. Prima facie, it cannot. On the contrary, according to this analysis deterministic and probabilistic laws supervene not only on the totality of particular facts, but also on the measures for simplicity, for strength, and for the goodness of balance; and these measures are something *we* add (at least as far as simplicity and balance is concerned; strength has at least an objective partial order).

Surely, Lewis cannot be on good terms with this apparent consequence of the best-system analysis. He shies away from any idealistic tendency like the devil from the holy water, also in order to maintain Humean Supervenience. However, he sees a way out: Perhaps nature is kind to us, and "if nature is kind, the best system will be *robustly* best – so far ahead of its rivals that it will come out first under any standards of simplicity and strength and balance" (Lewis 1994, p. 479 – his italics). If so, laws and chances do not depend on our inductive standards.

Yet, can there be a system that is robustly best under any standards? I guess even the kindest world is susceptible to transmogrification under gruesome standards. We may refer to factual human standards, but even there we find a lot of madness. Presumably, Lewis intends to quantify only over all reasonable inductive standards, and perhaps nature then has a better chance to be kind. Look, though, how wide the disagreement about reasonableness is, e.g., from the optimistic middle Carnap who had hoped for *the* inductive logic to the pessimistic subjectivists who plead for coherence and nothing more. It is quite obscure what a kind world is and how many of them there are.

In any case, Humean Supervenience turns out doubly constrained. It is ontologically restricted to worlds like ours devoid of alien matters, and it is epistemologically restricted to kind worlds free of indeterminateness concerning the best system. The two restrictions appear to be independent, and together they turn Humean Supervenience into an uncomfortable doctrine. I think that the problems in the last section about undermining futures constitute a telling objection. On the whole, the doctrine seems in need of getting straightened out.

As to the second critical point concerning objectivity and independence of our standards Lewis had envisaged another solution:

> I used to think rigidification came to the rescue: in talking about what the laws would be if we changed our thinking, we use not our hypothetical new standards of simplicity and strength and balance, but rather our actual and present standards. (Lewis 1994, p. 479)

Yes precisely. rigidification is one salient strategy of objectification.[27] Alas, Lewis continues:

> But now I think that is a cosmetic remedy only. It doesn't make the problem go away, it only makes it harder to state. (Lewis 1994, p. 479)

I did not understand this remark, so I requested him for clarification. Since I did not find the point explained elsewhere in his writings, let me quote extensively from his personal communication of February 13, 1996:

> Let me answer not your question but a generalization of it. The problem is that a certain analysis says that X (in this case, lawhood) depends on Y (in this case, our standards of simplicity, etc.) and yet we would ordinarily think this wasn't so. If Y were different, X would be just the same – or so we offhand think.
>
> A proposed answer is that 'X' is a rigidified designator of the actual value of something that depends on Y, and of course it's not true that the *actual* value would be different if Y were different. That's supposed to explain our opinion that there's no dependence.
>
> Well, *if* that's so – I'd think that it well might be so under at least some legitimate disambiguation – let '$\dagger X$' be a *derigidification* of the rigidified term 'X'. Maybe there's some nice ordinary-language reading of the derigidifying modifier; or maybe not, but in any case we can introduce it into our language by a suitable semantic explanation (as is done, for instance, in Stalnaker's paper 'Assertion', *Syntax and Semantics* 9).[28] Then it might turn out that our original opinion that X doesn't depend on Y survives in modified form: as the opinion that even $\dagger X$ doesn't depend on Y. If so, the alleged rigidification of X ends up making no difference. I think that's what does happen in the case of lawhood and our standards of simplicity etc. And that's why the hypothesis of rigidification, even if true, doesn't make the problem of counter-intuitive dependence go away. It makes it harder to state, because to state it you must first introduce the notion of derigidification.

He did not further explain, however, why the intuition that lawhood is independent of our standards should be maintained under derigidification. Projectivism, which I am going to recommend, does not share this intuition. The projectivist rigidifies the result of his projection and thus legitimately claims objectivity for this result. But he is content with so much objectivity. He would immediately grant that derigidification brings the process of projection back into focus and thus displays the dependence on the cognitive subject. However, there is no need to decide the dispute about intuitions. The point rather is that Lewis' idea which was not good enough for himself helps projectivism to some arguably sufficient notion of objectivity while allowing to admit, in another sense, the dependence of lawhood on our inductive standards.[29]

Projection Turns the Principal Principle into a Special Case of the Reflection Principle

The last remark puts the cart before the horse. We still do not know what the projectivistic understanding of chances is actually supposed to be. In order to explain it, let us follow the Lewisian track of the best-system analysis, but let us avoid, contra Lewis, to give it an ontological turn, let us rather keep it within its epistemological home. This will lead us onto well-trodden paths, but I said right at the beginning that there are no new discoveries to be made.[30]

The best system is, first of all, based on complete experience, on complete knowledge of particular observable facts. If these should be only finitely many, then all statistical methodology tells us that they do not allow for guaranteed conclusions with respect to objective probabilities; to force a decision, for whatever reasons, is simply unjustified. This conclusion certainly remains true when we include the broader inductive considerations relevant to best systems. If the set of particular facts should be infinite, the situation is not really different. If a die is actually cast infinitely many times, the propensities of the throws will change, simply because the die will physically change, and then the limit of relative frequency does not help us to a definite conclusion concerning the propensities. This is our epistemic situation vis à vis a small die, and I do not see why it should be different with respect to large worlds. In the strictest sense, nothing is repeatable. In saying this I flatly deny Humean Supervenience, of course.

Hence, it is actually unfeasible to precisely detect chances, even given complete knowledge of particular facts. The detectibility is rather merely counterfactual. Suppose we could run our world over and over again, indeed infinitely many times, suppose that all repetitions were governed by the same objective chance mechanism, and suppose we could learn all particular facts within not only one, but all repetitions. Then we would finally have established the chance law P_w of w, at least with probabilistic certainty. The last proviso is essential. If we live in a chancy world, we know a priori that there is a chance for misleading evidence, and we know a priori that even counterfactually ideal evidence cannot close the gap; the difference between probability 0 and impossibility is ineliminable.

If we want to describe this ideal detectibility of chances more formally, we obviously have to consider $W_0 \times W^\infty$, i.e., not only the original space $W = W_0$ of worlds of particular facts, but besides the space W^∞ of infinitely many possible counterfactual runs of the actual world; each $w^\infty \in W^\infty$ thus is an infinite sequence of possible worlds, each being a complete course of particular facts. (The term "W_0" is introduced only in order to distinguish that copy of W from its infinitely many counterfactual repetitions.) And we have to extend our probabilistic notions to $W_0 \times W^\infty$. If the actual world w is governed by the chance law $P_w \in \mathcal{P}$ defined for propositions over W_0, then these infinite sequences are governed by the product (or Bernoulli) measure $P_w^\infty \in \mathcal{P}^\infty$ which is the infinite product of P_w with itself and which is defined for propositions over W^∞. According to P_w^∞ the individual runs are governed by the same chance law P_w, and they are stochastically independent from one another; thus are our counterfactual suppositions for the ideal detectibility of P_w. Finally, we have to assume an a priori credence C_0^∞ also defined for propositions over $W_0 \times W^\infty$. C_0^∞ is not concerned with chances; it only captures our a priori expectations about all the particular facts in $W_0 \times W^\infty$. Of course, it extends the factual part of C_0; i.e., for each proposition $A \subseteq W_0$ we have $C_0^\infty(A) = C_0(A)$. I shall soon say a bit more about C_0^∞.

What we just said about the counterfactual detectibility of chances then condenses into what I would like to call the *Knowability Principle*:

(KP) $C_0^\infty(A|w^\infty) = P_w(A)$ P_w^∞-almost surely for all P_w and all $A \subseteq W_0$.

The left-hand side is indeed a random variable with w^∞ as random argument. That the equation holds P_w^∞-almost surely is to say that the set of w^∞ for which the equation holds has P_w^∞-probability 1. The expression "C_0^∞ $(A|w^\infty)$" is once more sloppy mathematics; it is short for the limit of the conditional credence of A when the condition infinitely grows into w^∞.

Instead of an ontologically conceived Humean Supervenience of chances on the actual particular facts, we thus have (KP) asserting the counterfactual knowability on the basis of counterfactual particular facts. We shall soon see how (KP) reduces to still more basic rationality constraints on C_0^∞. "Knowability" is perhaps too strong a word; strictly speaking, we can never know the chances, we can only be almost sure of them. However, (KP) captures all what (counterfactual) particular facts can tell us about chances; even counterfactually there is no more to know; (KP) is our best approximation to knowability.

I introduced (KP) only as our epistemological substitute for the misguided ontological Humean Supervenience of chances. In fact, (KP) follows from standard principles. So far, we have not yet explicitly considered relative frequencies. This is easily done, though. Let $rf(A)(w^\infty)$ stand for the limit (if it exists) of the relative frequency of the realization of the proposition A in the infinite random sequence w^∞. Then two further principles hold, namely the (strong) Law of Large Numbers:

$$\text{(LLN)} \quad rf(A)(w^\infty) = P_w(A) \ P_w^\infty\text{-almost surely for all } P_w \text{ and all } A \subseteq W_0,$$

and the so-called Reichenbach Axiom (recommended by Hilary Putnam to Carnap in 1953; cf. Carnap 1980, p. 120):

$$\text{(RA)} \quad C_0^\infty(A|w^\infty) = rf(A)(w^\infty) \text{ for all } w^\infty \text{ and all } A \subseteq W_0,$$

which says that our beliefs should increasingly and in the limit perfectly align with the observed relative frequencies, whatever they are. (KP), (LLN), and (RA) form a triangle connecting credence, chance, and relative frequency. Among the three, (LLN) and (RA) are the more basic ones. (LLN) is not a rationality postulate, but a mathematical theorem. Moreover, given (LLN), (KP) obviously follows from (RA), but not vice versa, because the equality of (KP) holds only almost surely.

Indeed, I find that de Finetti's representation theorem fits perfectly to my counterfactual set-up, thus providing further insight into the Reichenbach Axiom. This is why I have emphasized at the beginning of this paper that I do hardly more than rearrange de Finetti's philosophy of probability. The a priori credence C_0^∞ should be a symmetric measure over the product space, i.e., the event that n given propositions realize in the first n repetitions has the same credence as the event that these propositions realize in any other n repetitions. This seems even more compelling in our counterfactual set-up, where all repetitions are equal by fiat, than in any factual set-up. De Finetti's representation theorem tells that all and only symmetric measures are mixtures of product or Bernoulli measures, indeed unique mixtures. Hence, symmetry entails the principle of non-negative instantial relevance (cf. Humburg 1971, p. 228). Moreover, given symmetry, (RA) is equivalent to the assumption that the support or carrier of the mixture is the space of all product measures. This in turn makes clear that, given symmetry, (RA) entails the principle of positive instantial relevance (cf. Humburg 1971, p. 233). This may suffice as a brief reminder of the familiar epistemological home of the Reichenbach Axiom and thus of the epistemological grounds of the Knowability Principle.

My next point is that (KP) entitles us to project the credence C_0^∞ for $W_0 \times W^\infty$, i.e., for the actual world and its infinitely many counterfactual repetitions onto the credence C_0 for $W_0 \times \mathcal{P}$, i.e., for the actual world and its chance measure. The *Projection Rule* tells for each proposition $A \subseteq W_0$ and each set $\mathcal{Q} \subseteq \mathcal{P}$ of chance measures for W that:

$$\text{(PROJ)} \quad C_0(A \times \mathcal{Q}) = C_0^\infty(A \times \{w^\infty | C_0^\infty(.|w^\infty) \in \mathcal{Q}\}).$$

The Projection Rule thus says that a priori our credence that the true chance measure is in $\mathcal{Q}$ (and that some factual proposition A holds) is the same as our credence (that A holds and) that the counterfactual infinite evidence w^∞ moves us into some state in $\mathcal{Q}$.

Why is (PROJ) legitimate? (KP) says that for each possible P_w^∞ the set of w^∞ making C_0^∞ diverge from P_w is a P_w^∞-null set. Due to its symmetry, however, C_0^∞ is a mixture of all the P_w^∞. Hence, the set of w^∞ making C_0^∞ diverge from all measures in $\mathcal{Q}$ is also a C_0^∞-null set, because its C_0^∞-probability is a mixture of all the P_w^∞-null sets involved. Note, again, that (PROJ) is not an ontological thesis reducing chance to counterfactual infinite sequences of factual worlds. The ontological slack between truth and evidence is ineliminable. However, the ontological slack has not the slightest epistemological weight and cannot surface in the epistemological rule (PROJ); it is a genuine 'don't care'.

The upshot of these considerations is that the Minimal Principle is an immediate consequence of the Projection Rule. Take $\mathcal{Q} = \{P_w | P_w(A) = x\}$. Then (PROJ) specializes to

$$C_0(A | P_w(A) = x) = C_0^\infty(A | \{w^\infty | C_0^\infty(A | w^\infty) = x\}) = x.$$

And this is nothing but (MP*), which we have seen is all we need together with (RED) (and (AdP)) to duplicate Lewis' account. Thus, the replacement of the ontological doctrine of Humean supervenience by the epistemological Knowability Principle (which backed up the Projection Rule) at the same time replaces the conflict with (OP*) by a confirmation of (OP*).[31]

I find it illuminating to cast the point into a somewhat different form. For this purpose, we have to introduce the final player of my scenario, van Fraassen's so-called Reflection Principle. It is entirely about subjective probability. There we have static rationality postulates like Coherence or the axioms of mathematical probability, Regularity, Symmetry, etc., and we have dynamic rationality postulates the best known of which is, of course, the Rule of Conditionalization. About the most basic of these dynamic postulates is the *Reflection Principle*[32]:

$$\text{(RP)} \quad C_t(A | C_{t'}(A) = x) = x.$$

Here, C_t is the subject's credence or subjective probability at time t, and it is understood that t' is later than t. In other words, C_t specifies the prior and $C_{t'}$ the posterior probabilities of the subject. The Reflection Principle thus says: Given the condition that my future probability for some proposition is x, my present probability for it is also x. In short: I trust now what I assume to be my future belief.

It is clear why (RP) is called an auto-epistemic principle; it assumes that my future beliefs are the objects of my present beliefs (even only as a supposition). If one accepts the richer auto-epistemic framework, then (RP) proves to be a most general dynamic doxastic law entailing conditionalization and its generalizations; it is even amenable to a Dutch Book justification (cf. Gaifman 1988; Hild 1998b). It is also obvious that (RP) is a rationality postulate of restricted validity. For instance, I should not now trust my future beliefs I will have when drunken, and when now reading the newspaper I should believe (within limits) what I have read even given that tomorrow I will have forgotten what I have read. Hence, I should reasonably trust only those of my future beliefs that I have acquired in a reasonable fashion and that I entertain from a *superior* point of view, which is certainly provided by experience (and maybe in other ways as well).

The similarity between the Minimal and the Reflection Principle strikes the eye, though they are about different subject matters. However, the similarity is easily turned into entailment. Take (RP), replace C_t by the 'first' a priori credence C_0^∞ and $C_{t'}$ by the 'last' credence $C_0^\infty(.|w^\infty)$ counterfactually completely informed. (RP) thus spezializes to

$$(\mathrm{RP}^\infty) \quad C_0^\infty(A|C_0^\infty(A|w^\infty) = x) = x.$$

Note that (RP^∞) is in fact a theorem, not merely a rationality postulate. As above, (PROJ) finally turns (RP^∞) into (MP*).[33] To summarize, in counterfactual 'future' we are completely informed about the counterfactual manifestations of the propensities in w of particular facts, thus completely informed we can infer the chances in w, and hence (MP*) turns out as a special case of (RP).

Humean Projection

What is the significance of these mathematically trivial transformations? If projectivism is the doctrine that some objective traits of the world can only be understood as objectified projections of human attitudes, how does the previous section support projectivism concerning chances? To resume, the story is as follows: We postulate chances, and we know that they are different from our subjective probabilities. Yet, we also know the rational shape of our credences, we know how we change and improve them, we know according to (KP) that we cannot say anything better than that the chances are what our credences would be after that infinite counterfactual information, not by necessity, but with probability 1, and we know according to (PROJ) that we may identify our credences about chances with our credences about that counterfactual information and what we learn from it. We are aware of the ontological gap between chance and credence, but our epistemological bridge over it leaves nothing to be desired. In this sense I take chance to be a projection from credence.

Jeffrey (1965, Section 12.7) discusses the general idea that objective probabilities are *objectified* subjective ones, and in (2004, p. 19) he says, referring to Hume, that "chances are simply projections of *robust* features of judgmental probabilities from our minds out into the world" (his emphasis). Maybe he had the same picture in mind as the one developed here. However, objectification as he describes it in Jeffrey (1965, Section 12.7) is admittedly not very objective; it just means conditioning subjective probabilities wrt the true member of some partition of the possibility space considered (or the limit of these conditionings wrt a sequence of the true members of ever finer partitions). Of course, the result depends on the initial subjective probabilities as well as on the chosen partition. Jeffrey argues that this latitude has some advantages, but it seems clear that the general idea needs refinement.

Lewis (1980, pp. 278f.) is pleased that his account may be understood as offering such a refinement. According to (OP), it is the history-chance partition, as he calls it, which is *the* correct objectifying partition, and according to (OP*) it is more simply the chance partition consisting of all T_w themselves. Skyrms (1980, Section IA4) makes, in effect, the same proposal, though he opts for more pragmatic flexibility than Lewis and rather hides the chance nature of his conditioning partition. It is a matter of taste whether one should call this a confirmation or a trivialization of Jeffrey's general idea. In any case, Jeffrey (2004, p. 20) reminds us that "on the Humean view it is *ordinary* conditions, making no use of the word 'chance,' that appear" in the condition of (MP) or in the conditioning partition (my emphasis). Jeffrey insists on the point because otherwise his objectification idea has no prospect of offering an analysis of chance, a prospect Lewis (1980, pp. 288ff.) explicitly denies.

So, how does Jeffrey's general idea fare with Humean Projection as construed here? According to (PROJ) it is indeed the partition consisting of all T_w which is invoked in objectification; it is, however, to be conceived as the partition into all $\{w^\infty | C_0^\infty(.|w^\infty) = P_w\}$. Hence, we have obeyed Jeffrey's reminder; we have used ordinary conditions making no use of the word "chance". Still, I am not sure whether Jeffrey would be satisfied. His examples always use partitions of the original possibility space W of particular facts, whereas I move to a partition of the possibility space W^∞ of infinite counterfactual repetitions of W. Only there particular facts can get as close to objective chances as they can get; and if this is so, then Jeffrey's objectification within the space W can at best reach pragmatically weakened forms. The detour via W^∞ appears unavoidable to me.

My continuous massive invocation of counterfactuality may have raised, however, suspicions from the outset. Skyrms (1980, p. 31) has already warned that "attempts to construe propensities as modalized relative frequencies *only make things worse* in this regard" (his emphasis), the regard being the use of the law of large numbers as an analysis of propensity. Skyrms is right. We have seen that chances do not ontologically reduce even to propositions over the counterfactual space W^∞; the slack *is* ineliminable. However, W^∞ serves here only epistemological, not ontological purposes.

For the same reason I am not worried by Lewis (1994, p. 477), when he says "I think that's a blind alley", thereby referring to "thinking of frequencies not in our actual world, but rather in counterfactual situations" (in order to deal with his puzzling case of unobtainium). Within his set-up he is indeed right. There, relative frequencies in counterfactual situations can inform us about the actual world, only if we have ascertained beforehand that the counterfactual situation is governed by the same chance law as the actual world. Thus, we would have to solve, according to Lewis, the supervenience issue for the counterfactual situation in order to solve it for the actual world; and this merely defers the issue. However, this is not our problem. We do not have the telescope view onto counterfactual situations, to use Kripke's terms; it was rather part of our counterfactual stipulation that all repetitions of W be governed by the same chance law P_w; there is no need to *ascertain* the chance law of the repetitions. I do not see why this counterfactual stipulation should be illegitimate. We always think about counterfactual situations and what we would believe given this or that situation, and in order to get Humean Projection running in our way, we only consider extreme cases of this kind. Specializing, or extending, (RP) to (RP^∞), in order to derive (MP), is not a misuse of the Reflection Principle; it is an extreme, though legitimate use.

Well, it may be legitimate; still it hardly helps. Given the extreme counterfactual evidence we may be as certain about chances as we can. Our actual evidence, however, is infinitely poorer. Indeed doubly so; we can inquire only a tiny part of our actual world and never the counterfactual repetitions. The counterfactual construction may, and should, I think, satisfy philosophers, but it is of no use for scientists and statisticians who cannot do better than gathering actual evidence and drawing conclusions from this insufficient basis. This, however, is something to acknowledge, not to deplore. The philosophical account provides the ideal standard, and it then is a methodological issue how best to approximate the ideal within our factual limits. Statisticians have developed most sophisticated test methods, of which randomization is an important part. But there are also more general preconceptions:

In principle, the scheme of partial determination governing our world may be any chance measure whatsoever. In principle, the whole world has the propensity to move into this or that state, and propensities may vary from here to there and from now to then. In our counterfactual scenario we could discover any wild distribution of chances, but in the actual world we want to understand the 'mechanics' of partial determination. The ground rule is: equal causes, equal effects; or rather, equal conditions, equal propensities – which gets bite only by restricting "equal". The relevant conditions should be few, not many. If we are lucky, we have kept constant all relevant conditions during a row of some thousands throws of some die, and then we may take the actual row as approximating the counterfactual sequence. The relevant conditions should be local, or contiguous, to use Hume's term. Non-locality is one of the mysteries created by quantum mechanics. Crystal balls are miraculous for the same reason. I find it incoherent to say that a given type of events is only partially determined, but can be unfailingly foreseen with a certain crystal ball. Rather, I would then take these events as fully determined – but would not understand how determination, i.e., the crystal ball works in these cases. If we are lucky, we shall be able to construe the chance law governing our world as a Markov process. If we develop different ideas about space and time, we have to adapt our preconceptions of the 'mechanics' of determination. And so forth.

If we do not succeed with our preconceptions, it is unclear how we would respond. In the extreme case, the idea of partial (or full) determination would dissolve. Thus it seems obvious to me that there is more to the notion of chance than just the Principal Principle. There are also all these preconceptions connecting chance with space and time, simplicity, orderliness, and whatnot. It is such things mentioned by Arntzenius, Hall (2003, pp. 177f.) when they arrive at the same conclusion. These preconceptions are modifiable, but only within limits; beyond the notion of chance will crumble.

Do such considerations reintroduce the epistemological double standard of which I have accused Lewis in the section *Humean Supervenience*? No. With regard to the ideal counterfactual evidence we can simply stick to de Finetti's story of the symmetric a priori credence satisfying the Reichenbach axiom and thus converging almost surely to the true chance measure, whatever it may be. Here, we do not need help from the additional considerations just mentioned. We have to rely on them when and because we try to make sense of our very restricted evidence. Thus, the second epistemological story that I have just indicated does not interfere with, but rather complements, the account I have extensively presented.

Since we have sacrificed Humean Supervenience, we also have avoided the ontological double standard and the resulting conflict between (OP) and (NP). We can, and do, simply stick to (OP*) and reject (NP*) as nonsense.

However, if we sacrifice (HS), we cannot do so without considering Lewis' two main reasons for it. The one consists in his ontological preferences. Without doubt, if (HS) were true, the resulting ontological picture would be most elegant and satisfying. Those rejecting Humean supervenience have different preferences and acknowledge irreducible dispositions, capacities, causes, necessities, or propensities. The projectivist, in particular, has a special story to tell about these matters that explains them ultimately with our subjective condition without diminishing their objectivity. I do not think that this ontological dispute can be resolved with general arguments. It is a matter of details, and there we have at least seen that Lewis had difficulties to maintain his prima facie elegance.

His second reason, though, is more pertinent and more urgent. It is best put in Lewis (1986, pp. xvf.):

> I could admit that . . . the chances . . . do not supervene on the arrangement of qualities. . . . Why not? I am not moved just by loyalty to my previous opinions. That answer works no better than the others. Here again the unHumean candidate for the job turns out to be unfit for its work. The distinctive thing about chances is their place in the 'Principal Principle,' which compellingly demands that we conform our credences about outcomes to our credences about chances. . . . I haven't the faintest notion how it might be rational to conform my credences about outcomes to my credences about some mysterious unHumean magnitude. Don't try to take the mystery away by saying that this unHumean magnitude is none other than *chance*! I say that I haven't the faintest notion how an unHumean magnitude can possibly do what it must do to deserve the name – namely, fit into the principle about rationality of credences – so don't just stipulate that it bears that name. Don't say: here is chance, now is it Humean or not? Ask: . . . Is there any way that an unHumean magnitude could [fill the chance-role]? . . . the answer is 'no'. . .

He repeats the point in Lewis (1994, pp. 484f.) with more confidence, having been shown a way out of the paradox of undermining futures generated by (HS).

His own response to this challenge is (HS). It is no mystery how particular facts constrain credence; and if chance supervenes on particular facts, it is in principle no mystery how chance constrains credence. And thus he sets out to remove paradox by modifying (OP). Right at the beginning of the section *Chance-Credence Principles* I indicated that this is the basic puzzle affecting the Principal Principle. The quotation indeed suggests that Lewis thinks that (HS) is the *only* solution of the puzzle (even though his challenge is directed foremost to the position of David Armstrong). How may the projectivist respond?

For the projectivist the puzzle has a straightforward solution.[34] This is clear from his general strategy. For him, chances are not alien features cognitive access to which is bound to be mysterious; they are of our own breeding. We need not speak figuratively, though; we have prepared a precise answer. Lewis is right; there is no mystery how particular facts constrain credence. However, van Fraassen is also right; there is in principle no mystery how future credence can constrain present credence. And we have seen that according to the projectivistic reconstrual the Principal Principle is nothing but an extreme application of the Reflection Principle. This was the whole point of my construction in the previous section. To be sure, in that application a priori credence is constrained by an extremely counterfactual 'future' credence. However, it is mostly counterfactual future credence to which (RP) applies, and we should certainly not bother about being more or less extreme. In this way, the projectivist is able to remove the puzzling air from the Principal Principle. Chance, being almost surely identical to projected credence objectified, must constrain a priori credence precisely in the way summarized in (OP*).

Appendix on Ranking Functions and Deterministic Laws: The Same All Over Again

The whole of this paper immediately and perfectly carries over to full determination or natural necessity and deterministic laws. Lewis tells the same story, this story meets the same criticism, and I have a precise projectivistic substitute story. Indeed, all this is more or less a matter of routine; I do not have to write a twin paper. Let me just indicate the basic points.

A very common, and also Lewis', assumption is that laws are regularities which in turn are mere generalizations expressed by universally quantified sentences. However, not all regularities are laws; we have to be selective. Lewis offers his best-system analysis of laws in order to discriminate them from mere regularities. He thinks laws Humeanly supervene on particular facts, and he constrains the supervenience of laws in the same way as that of chance. The only point missing is that the Principal Principle and the ensuing discussion have no explicit deterministic counterpart.

The problems remain. (HS) is again ontologically as well as epistemologically constrained. Carroll (1994, Chapter 3) and Ward (2002, Section 3) attempt to specify examples of two worlds in which the same facts, but different laws obtain. Black

(1998, p. 376) suggests "that laws can ... undermine themselves, in that the laws of the universe might allow that the laws of the universe could have been otherwise." Hence, it looks like we are running into the same kind of problems with deterministic laws as we have extensively discussed here with respect to chance. Lewis' reasons for sticking to (HS) are also the same. Without (HS) we could not understand the idea of necessitation. Hence, the dialectic situation is as before. What, though, could be the constructive alternative? This is indeed much less clear than in the probabilistic case where subjective and objective probabilities and their delicate relation are perhaps not fully understood, but familiar for a very long time.

I think the basic mistake lies already in the common assumption that laws are (a special kind of) regularities. In this respect, laws are much more deceptive than chances. One immediately sees that chances are modalities; they take propositions as arguments and somehow assign numbers to them. By contrast, laws appear to be mere propositions, and modality is prima facie not involved. Any subsequent mounting of modality is then bound to create mysteries. The alternative, though, is not to start with a primitive necessitation operator, as Armstrong does in his analysis of lawhood. This is no less mysterious. Also, it will not do to conceive of deterministic laws as a limiting case of chance laws, not only for the reason that a chance of 1 is not quite necessitation. This is not the place, however, to go through all the various accounts of lawhood. Let me just say that I believe that the alternative must be somehow to tell the same kind of story as we did in the probabilistic case. But how?

The answer is: with the help of ranking functions (first presented in Spohn (1983, 1988), where I called them ordinal conditional functions). As in probability, we must start with the subjective side, with the representation of belief. This is what a ranking function does. A *ranking function* κ for a given possibility space W is a function from W into the set of nonnegative integers such that $\kappa(w) = 0$ for some $w \in W$. The ranking is extended to propositions $A \subseteq W$ by defining $\kappa(A) = \min\{\kappa(w)|w \in A\}$. And conditional ranks are defined by $\kappa(B|A) = \kappa(B \cap A) - \kappa(A)$.

Ranks are degrees of disbelief. $\kappa(A) = 0$ says that A is not disbelieved at all; $\kappa(A) = n > 0$ says that A is disbelieved to degree n. Hence, $\kappa(\overline{A}) > 0$ expresses that $\overline{A}$ is disbelieved (to some degree) and hence that A is believed (to the same degree). Thus, ranking functions, unlike probability measures, represent belief (acceptance, holding to be true). This is their most distinctive feature due to which they can be related to deterministic as opposed to probabilistic laws. Unlike doxastic logic, or even AGM belief revision theory, ranking functions can also account for a full dynamics of belief; this means at the same time that they embody a full inductive logic. Basically, this dynamics consists in conditionalization, just as in probability theory.[35] The reason why this works perfectly is that conditional ranks as defined above behave almost exactly like conditional probabilities. Indeed, the parallel extends much farther. Practically all virtues of Bayesian epistemology can be carried over to ranking functions. (For a fuller explanation of these claims see Spohn 1988, 2009.)

One thing we can now do, for instance, is to state the Reflection Principle in ranking terms:

$$(\text{RP}\kappa) \quad \kappa_t(A|\kappa_{t'}(A) = n) = n,$$

which says that given you disbelieve A tomorrow to degree n you so disbelieve it already today. (RPκ) is indeed a strengthening of Binkley's principle.[36] All the remarks about the probabilistic version (RP) apply here as well.

I have emphasized that ranking functions must be interpreted as representing doxastic states. They represent what a subject takes to be true or false, but they are not true or false themselves. However, to some extent they can be objectified so that it makes sense to apply truth and falsity to them, just as to propositions. How this objectification works is a somewhat tricky story elaborated in Spohn (1993). According to this story, most ranking functions cannot be objectified. This appears to be different with chances. Any credence measure for W could, it seems, also serve as a chance measure for W. But maybe not. We have seen above that there is more to chance and that one might, for instance, suggest that only probability measures representing a Markov process can be chance measures.

Anyway, what I have proposed in Spohn (1993) is that causal laws or, in the present terms, schemes of full determination are just such objectifiable ranking functions, a view I have philosophically more thoroughly explained in Spohn (2002). The crucial point is that the inductive behavior is thus directly built in into laws and not subsequently imposed on something propositional. Moreover, for laws so conceived we can tell de Finetti's complete story as shown in detail in Spohn (2005): If the ranking function κ is such a scheme of full determination for W, we can again form the infinite product space W^∞ and the product ranking function κ^∞ independently repeating κ infinitely many times. Any symmetric ranking function over W^∞ is then a unique mixture of such product ranking functions, which will converge to the true law (= product ranking function) with increasing evidence. Hereby, the role relative frequencies have in the probabilistic case is taken over by the number of exceptions in the deterministic case.

In sum, we have here all the ingredients for telling exactly the parallel story about necessitation or full determination as we have told about partial determination. Deterministic laws are, in the way explained, projections of ranking functions, i.e., of subjective states representing beliefs and their dynamics.

Acknowledgements I am most grateful to Ludwig Fahrbach and Jacob Rosenthal for thoroughgoing discussions of earlier drafts of this paper; it gained immensely thereby.

Notes

[1] I have written a minor note, Spohn (1987), which foreshadows the general line of thought, and a German paper Spohn (1999), of which the present paper is a substantial elaboration.

[2] There presumably are deep connections between metaphysical and natural necessity. Still, the two kinds of necessity must at first be kept apart. Metaphysical necessity is tied up with identity and existence, natural necessity is not, prima facie. Here, I shall deal only with the latter without worrying about its connection to the former.

[3] Similar phrasings may be found in Popper (1990, pp. 18f.) and Miller (1995, p. 138).

[4] For instance, Fetzer (2002) shares realism about propensities, but responds to such concerns by embedding propensities into an embracive account of explanation and abductive inference.

While I am in sympathy to his general approach, I do not want to explicitly enter the topic of explanation. Of course, that topic is tightly interwoven with our present one, but it has its own intricacies, in particular, when it comes to saying what 'the best explanation' might be. As far as I can see, we shall be able to side-step these intricacies here without loss.

[5] My reference book is Rosenthal (2004) that offers forceful criticisms of prominent variants of the propensity interpretation.

[6] Logue (1995) apparently pursues the same goal. However, he insists on having only one notion of probability, a personalistic one, and he does not present an explicit projectivistic construal of objective probability. The only further probability book where the idea is taken up is Rosenthal (2004, pp. 199ff.). In fact, the challenge of understanding objective probability as it is built up in this book in a most pressing way provoked me to elaborate my (1999) into the present paper.

[7] Concerning deterministic laws, Ward (2002) also claims to give a projectivistic account which he extends to chance in Ward (2005). However, while I agree with his critical diagnosis, our constructive approaches widely diverge, as will become clear at the end of this paper.

[8] Often, direct inference is more narrowly understood as the more contested 'straight rule' that recommends credence to equal observed relative frequency.

[9] The puzzle is vividly elaborated by Rosenthal (2004, Section 6.3).

[10] I shall even prefer sentential over set theoretical representations of propositions.

[11] This is Constraint 2 of Skyrms (1980, pp. 163–165), applied to degrees of belief and propensities.

[12] Even at the risk of appearing pedantic, let me at least once note what the correct set-theoretic representation of (MP*) is. There, credence is not about facts and chance, but rather about facts and evolutions of chance, i.e., about $W \times \mathcal{P}^T$, where T is the set of points of time. (Only at the end of the next section shall we be able to return to our initial simpler conception of credence.) (MP*) then says that $C_0(A|\{\pi \in \mathcal{P}^T | \pi(t)(A) = x\}) = x$, where the condition consists of all those evolutions of chance according to which the chance of A at t is x.

[13] Hall (2004, p. 103) arrives at another definition of admissibility. However, it is clear that his move from his (3.12) to his (3.13) offers another sufficient condition the necessity of which is not argued for. In any case, his definiens entails mine, but not vice versa.

It is unfortunate that my paper was essentially finished before I could get aware of this paper of Ned Hall, which covers much of the same ground as mine, though with different twists and conclusions. Thus, my comparative remarks will be confined to some footnotes.

[14] Of course, I am presupposing classical time throughout. I do not venture speculating about the consequences of relativistic time for our topic.

[15] Proof: According to (CP) we have $C_0(A|H_{wt} \;\&\; P_{wt}(A|H_{wt}) = x) = x$. According to (DP) $P_{wt}(A|H_{wt}) = x$ expresses the same proposition as $P_{wt}(A) = x$. Hence, we also have $C_0(A|H_{wt} \;\&\; P_{wt}(A) = x) = x$. This just says that H_{wt} is admissible wrt A in w at t. Since any t–historical E is the disjoint union of some H_{wt}, E is admissible, too, wrt A in w at t.

[16] For a recent reinforcement of the problem see Hájek (2007).

[17] This follows from the graphoid axioms for conditional probabilistic independence; cf., e.g., Spohn (1978, pp. 102f.).

[18] Why is P now double-indexed? Because we have to say for the world w not only what the chances are in w at t $(= P_{wt})$, but also what the chances would have been if H_{vt} had been its history up to t $(= P_{w,vt})$.

[19] This is different from the identification of probabilities of conditionals with conditional probabilities, of which Lewis (1976) has warned us.

[20] The satisfiability of the consistency requirement is obvious in the case of discrete time with a first point of time. In the other cases one has to allude to convergence theorems for descending martingales; cf., e.g., Bauer (1968, p. 281).

[21] Hall (2004, p. 96) undertakes the same reduction. P_w is what he calls ur-chance.

[22] If this algebra were a complete one, this translation of (HS) would indeed be correct.

[23] We might, of course, strengthen (HS) to the effect that chances at t supervene on no more than factual history up to t; then chance information is only about the past, and the paradox cannot arise. Lewis (1980, pp. 291f.) already mentions this option before clearly seeing the paradox. In 1986 (p. 131) he even expresses a preference for it after recognizing the paradox. In 1994 (Section 6) he finally rejects it, rightly in my view.

[24] Proof: Take (CP), specialize B to H_{wt} & T_w and apply (DP) for omitting H_{wt} from the condition of P_{wt}. Then you get $C_0(A|H_{wt} \,\&\, T_w \,\&\, P_{wt}(A|T_w) = x) = x$, which is nothing but (NP), since H_{wt} & T_w entail $P_{wt}(A|T_w) = x$.

[25] Cf. Arntzenius, Hall (2003, pp. 176f.) for a fuller explanation of the point.

[26] Sturgeon (1998) argues that the restrictions put on (HS) are indeed incoherent, however they are specified. Hall (2004, pp. 108ff.) also critically discusses how the inference of chances from facts is supposed to go.

[27] Loewer (1996, pp. 114f.) also discusses the point and recommends rigidification.

[28] This refers to Stalnaker (1978).

[29] The point is indeed one of deep and general importance. It applies, I believe, to objecthood in general, certainly a most fundamental matter. *We* cut up the world into pieces, *we* constitute objects by saying which properties or kinds of properties are essential or constitutive for them. (This allows for the case that we fix only a space of possible essential properties of an object and leave it to the actual world to fix the actual essential properties.) Still, the objects thus constituted are objects independent of us, their objectivity is in no way impaired by our constituting them, in particular because we constitute objects in such a way that our constituting is *not* essential for them. The point extends to properties. In two-dimensional semantics each predicate expresses a (derigidified) concept and denotes a (rigidified) property, and while most concepts are, as I say, a priori relational, only few properties are necessarily relational – two notions of relationality that are particularly relevant vis à vis color predicates; cf. my (1997, pp. 367ff.). This footnote indicates the direction into which this paper would need most to be further thought through.

[30] This section elaborates the core of the predecessor paper Spohn (1999).

[31] Hall (2004, pp. 108f.) envisages the same kind of argument, also with reference to de Finetti's representation theorem, though without actually endorsing it. He ascribes the argument to a position he calls 'primitivist hypothetical frequentism, which, however, is not mine. As he describes it, this kind of frequentist equates chance with limiting hypothetical relative frequency and considers it to be a brute metaphysical fact that this equation is correct. By contrast, I emphasized the almost unnoticeable epistemological-ontological gap, and I do not see the necessity to close it per fiat.

[32] The Reflection Principle is explicitly stated in van Fraassen (1984); there its deep philosophical relevance was fully recognized. He returns to it at length in van Fraassen (1995). Other references are Goldstein (1983) and Spohn (1978, pp. 162f.) where I stated an equivalent principle (called the Iteration Principle by Hild 1998a, p. 329) within an auto-epistemic or reflexive decision-theoretic setting and under the restrictions usually accepted nowadays. Penetrating discussions may in particular be found in Hild (1998a, b).

[33] Skyrms (1980, Appendix 2) already observed that there is a common form to such principles that is open to various interpretations. Following Gaifman (1988), the common form might be called 'expert principle', since it describes trust in some kind of expert. For this unified view see in particular Hall (2004) and Hájek (2007). However, it is only our Projection Rule which establishes an entailment between the expert principles considered here, i.e., (RP) and (MP).

[34] As mentioned in footnote 31, Hall (2004, p. 109) also envisages the solution defended here (with some doubts concerning its general feasibility). However, he envisages it only as a possibility in order to prove the point he is up to in his paper, viz., that the reductionist claiming (HS) need not have an advantage over the non-reductionist vis à vis this issue. For him (cf. p. 107), a no less acceptable response seems to be to declare the Principal Principle analytic and to reject any further justificatory demands. As I have explained in the Section Chance-Credence Principles, this will not do. We have a real challenge here which requires some substantial response.

[35] The idea that belief is just probability 1 is not only intuitively unsatisfactory, but also theoretically defective, because conditionalization does not work for extreme probabilities and beliefs could then only be accumulated and never revised. (Popper measures solve this problem just as half-way as does AGM belief revision; see Spohn 1986.) This is the essential reason why it does not work to correspondingly conceive deterministic laws as limiting cases of chance laws.

[36] It says that if I believe now that I shall believe tomorrow that p, I should already now believe that p. Binkley (1968) introduced it in relation to the surprise examination paradox.

References

Armstrong DM (1983) What is a law of nature? Cambridge University Press, Cambridge
Armstrong DM (1997) A world of states of affairs. Cambridge University Press, Cambridge
Arntzenius F, Hall N (2003) On what we know about chance. Br J Philos Sci 54:171–179
Barwise J, Etchemendy J (1987) The liar: an essay on truth and circularity. Oxford University Press, Oxford
Bauer H (1968) Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie. de Gruyter, Berlin
Binkley RW (1968) The surprise examination in modal logic. J Philos 65:127–136
Black R (1998) Chance, credence, and the principal principle. Br J Philos Sci 49:371–385
Blackburn S (1993) Essays in quasi-realism. Oxford University Press, Oxford
Carnap R (1980) A basic system of inductive logic, part 2. In: Jeffrey RC (ed.) Studies in Inductive Logic and Probability, vol. II. University of California Press, Berkeley, pp 7–155
Carroll JW (1994) Laws of nature. Cambridge University Press, Cambridge
Cartwright N (1989) Nature's capacities and their measurement. Clarendon, Oxford
Chaitin GJ (1966) On the length of programs for computing finite binary sequences. J Assoc Comp Mach 13:547–569
Church A (1940) On the concept of a random sequence. Bull Am Math Soc 46:130–135
Fetzer JH (2002) Propensities and frequencies: inference to the best explanation. Synthese 132: 27–61
Gaifman H (1988) A theory of higher order probabilities. In: Skyrms B, Harper WL (eds) Causation, chance, and credence. Kluwer, Dordrecht, pp 191–219
Gillies D (2000) Philosophical theories of probability. Routledge, London
Goldstein M (1983) The prevision of a prevision. J Am Stat Assoc 78:817–819
Hájek A (2007) The reference class problem is your problem too. Synthese 156:185–215
Hall N (1994) Correcting the guide to objective chance. Mind 103:505–517
Hall N (2004) Two mistakes about credence and chance. In: Jackson L, Priest G (eds) Lewisian themes: the philosophy of David K. Lewis. Oxford University Press, Oxford, pp 94–112
Hild M (1998a) Auto-epistemology and updating. Philos Stud 92:321–361
Hild M (1998b) The coherence argument against conditionalization. Synthese 115:229–258
Hoefer C (1997) On Lewis's objective chance: 'humean supervenience debugged'. Mind 106: 321–334
Humburg J (1971) The principle of instantial relevance. In: Carnap R, Jeffrey RC (eds) Studies in inductive logic and probability, vol I. University of California Press, Berkeley, pp 225–233
Jeffrey RC (1965) The logic of decision. The University of Chicago Press, Chicago, 2nd ed. 1983
Jeffrey RC (2004) Subjective probability. The real thing. Cambridge University Press, Cambridge
Lewis D (1976) Probabilities of conditionals and conditional probabilities. Philos Rev 85:297–315
Lewis D (1980) A subjectivist's guide to objective chance. In: Jeffrey R (ed) Studies in inductive logic and probability, vol II. University of California Press, Berkeley, CA, pp 263–293; with postscripts also in Lewis (1986), pp 83–132
Lewis D (1986) Philosophical papers, vol II. Oxford University Press, Oxford
Lewis D (1994) Humean supervenience debugged. Mind 103:473–490; also in: Lewis D (ed) Papers in metaphysics and epistemology. Cambridge University Press, Cambridge 1999, pp 224–247
Loewer B (1996) Humean supervenience. Philos Topics 24:101–127
Logue J (1995) Projective probability. Oxford University Press, Oxford
Miller D (1966) A paradox of information. Br J Philos Sci 17:59–61
Miller D (1995) Propensities and indeterminism. In: O'Hear A (ed) Karl Popper: philosophy and problems. Cambridge University Press, Cambridge, pp 121–147
Popper KR (1990) A world of propensities. Thoemmes, Bristol
Rosenthal J (2004) Wahrscheinlichkeiten als Tendenzen. Eine Untersuchung objektiver Wahrscheinlichkeitsbegriffe. Mentis, Paderborn
Salmon WC (1984) Scientific explanation and the causal structure of the world. Princeton University Press, Princeton, NJ

Schaffer J (2003) Principled chances. Br J Philos Sci 54:27–41
Skyrms B (1980) Causal necessity. Yale University Press, New Haven, CT
Spohn W (1978) Grundlagen der Entscheidungstheorie. Scriptor, Kronberg (out of print, pdf-version at: http://www.uni-konstanz.de/FuF/Philo/Philosophie/Mitarbeiter/spohn.shtml)
Spohn W (1983) Eine Theorie der Kausalität. Unpublished Habilitationsschrift, University of Munich, pdf-version at: http://www.uni-konstanz.de/FuF/Philo/Philosophie/Spohn/spohn_files/Habilitation.pdf
Spohn W (1986) The representation of popper measures. Topoi 5:69–74
Spohn W (1987) A brief remark on the problem of interpreting probability objectively. Erkenntnis 26:329–334
Spohn W (1988) Ordinal conditional functions. A dynamic theory of epistemic states. In: Harper WL, Skyrms B (eds) Causation in decision, belief change, and statistics, vol II. Kluwer, Dordrecht, pp 105–134
Spohn W (1993) Causal laws are objectifications of inductive schemes. In: Dubucs J (ed) Philosophy of probability. Kluwer, Dordrecht, pp 223–252
Spohn W (1997) The character of color predicates: a materialist view. In: Künne W, Newen A, Anduschus M (eds) Direct reference, indexicality and propositional attitudes. CSLI Publications, Stanford, pp 351–379
Spohn W (1999) Lewis' principal principle ist ein Spezialfall von van Fraassens reflexion principle. In: Nida-Rümelin J (ed) Rationalität, Realismus, Revision. de Gruyter, Berlin, pp 164–173
Spohn W (2002) Laws, ceteris paribus conditions, and the dynamics of belief. Erkenntnis 57: 373–394; also in: Earman J, Glymour C, Mitchell S (eds) Ceteris paribus laws. Kluwer, Dordrecht, pp 97–118
Spohn W (2005) Enumerative induction and lawlikeness, Philos Sci 72, 164–187
Spohn W (2009) A Survey of Ranking Theory. In: Huber F, Schmidt-Petri C (eds) Degrees of belief. An anthology. Springer, Dordrecht, pp 185–228
Stalnaker RC (1978) Assertion. In: Cole P (ed) Syntax and semantics, vol 9. Pragmatics. Academic, New York, pp 315–332
Strevens M (1995) A close look at the 'new' principle. Br J Philos Sci 46:545–561
Sturgeon S (1998) Humean chance: five questions for David Lewis. Erkenntnis 49:321–335
Thau M (1994) Undermining and admissibility. Mind 103:491–503
van Fraassen BC (1980) A temporal framework for conditionals and chance. Philos Rev 89:91–108
van Fraassen BC (1984) Belief and the Will. J Philos 81:235–256
van Fraassen BC (1995) Belief and the problem of Ulysses and the Sirens. Philos Stud 77:7–37
von Mises R (1919) Grundlagen der Wahrscheinlichkeitsrechnung. Mathematische Zeitschrift 5:52–99
Vranas P (2004) Have your cake and eat it too: the old principal principle reconciled with the new. Philos Phenomenol Res 69:368–382
Ward B (2002) Humeanism without humean supervenience: a projectivist account of laws and possibilities. Philos Stud 107:191–218
Ward B (2005) Projecting chances: a humean vindication and justification of the principal principle. Philos Sci 72:241–261

Evolutionary Theory and the Reality of Macro-Probabilities

Elliott Sober

Evolutionary theory is awash with probabilities. For example, natural selection is said to occur when there is variation in fitness, and fitness is standardly decomposed into two components, viability and fertility, each of which is understood probabilistically. With respect to viability, a fertilized egg is said to have a certain *chance* of surviving to reproductive age; with respect to fertility, an adult is said to have an *expected number* of offspring.[1] There is more to evolutionary theory than the theory of natural selection, and here too one finds probabilistic concepts aplenty. When there is no selection, the theory of neutral evolution says that a gene's chance of eventually reaching fixation is 1/(2N), where N is the number of organisms in the generation of the diploid population to which the gene belongs. The evolutionary consequences of mutation are likewise conceptualized in terms of the probability per unit time a gene has of changing from one state to another. The examples just mentioned are all "forward-directed" probabilities; they describe the probability of later events, conditional on earlier events. However, evolutionary theory also uses "backwards probabilities" that describe the probability of a cause conditional on its effects; for example, coalescence theory allows one to calculate the expected number of generations in the past that the genes in the present generation find their most recent common ancestor.

If evolutionary *theory* is inundated with probabilities, is the same true of the *processes* that evolutionary theory seeks to characterize? A straightforward realist interpretation of the theory yields an affirmative answer to this question. Since the theory truly describes what happens in nature, and since the theory describes nature probabilistically, the probabilities it postulates are real. In spite of the simplicity of this interpretation, there have been dissenters. The title of Alexander Rosenberg's (1994) book, *Instrumental Biology or the Disunity of Science*, suggests where he stands on this issue. Rosenberg's thesis is that the probabilities used in evolutionary theory should *not* be interpreted realistically – they are not objective quantities – because they are mere excuses for our ignorance of detail. For Rosenberg, "evolutionary phenomena are . . . deterministic, or at least as deterministic as underlying quantum indeterminism will allow (p. 82)."[2] The probabilities

E. Sober (✉)
Philosophy Department, University of Wisconsin, Madison, Wisconsin 53706, USA
e-mail: ersober@wisc.edu

E. Eells and J.H. Fetzer (eds.), *The Place of Probability in Science*, Boston Studies in the Philosophy of Science 284, DOI 10.1007/978-90-481-3615-5_6,

of evolutionary *phenomena* are one thing, the probabilities that evolutionary *theory* assigns to those phenomena another.[3]

Although Rosenberg's thesis is about evolutionary theory, his reasons for holding it are general enough to apply to any theory that uses probabilities. And because the motivation for Rosenberg's instrumentalism is so general, it is no surprise that this position was enunciated long before evolutionary theory was mathematized in the twentieth century. Rosenberg's thesis traces back to Laplace.

> Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it – an intelligence sufficiently vast to submit these data to analysis – it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, and the past, would be present to its eyes (Laplace 1814, p. 4).

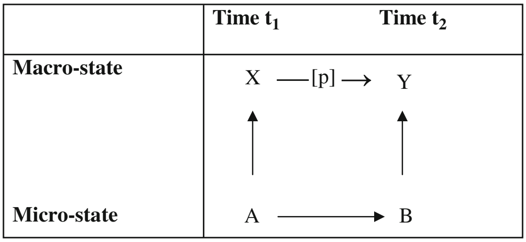
In the *Origin*, Darwin (1859, p. 131) gives voice to the same thought when he explains what he means by saying that variation is "... due to chance. This, of course, is a wholly incorrect expression, but it serves to acknowledge plainly our ignorance of the cause of each particular variation."

Laplace was thinking about Newtonian theory when he described his demon, and he took that theory to be deterministic in form.[4] Since that theory makes no use of probabilities, the probabilities we use to describe nature are mere confessions of ignorance. Rosenberg's Laplacean position involves no commitment to determinism. He concedes that if determinism is false, then it is a mistake to claim that the *only* reason we use probabilities to describe nature is that we are ignorant of relevant details. But there is a Laplacean thesis that survives the death of determinism; this is the reductionist idea that *the only objective probability an event has is the one assigned to it by the micro-theory*.[5] If Newtonian theory is the true theory of particles, then no probabilities (other than zero and one) are objective. If quantum mechanics is the true theory of particles, then the only objective probabilities are the ones assigned by quantum mechanics. Either way, the probabilities assigned by evolutionary theory are not to be interpreted realistically if they differ in value from the ones assigned by whatever the true micro-theory turns out to be.

It is interesting that Laplace says that his demon would find *nothing* uncertain; this goes beyond the more modest claim that the demon would have no uncertainty about the mass, velocity, acceleration, and other properties described by Newtonian theory. Laplace's stronger claim suggests the thought that the properties discussed in Newtonian theory provide a *synchronic supervenience base* for all the other properties that macro-objects might have. And this thought, in turn, can be generalized further, so that there is no reliance on Newtonian theory or on the truth of determinism:

(MS) A complete specification of the properties that all particles have at a given time uniquely determines *all* the properties that *all* macro-objects have at that same time.

Fig. 1 Laplace's demon in a deterministic Universe



This is the idea of *mereological* (part/whole) supervenience. If two objects and the environments they occupy are particle-for-particle physical copies of each other, they will also be identical in terms of their psychological and biological properties.[6]

The principle of mereological supervenience (MS) says that micro determines macro. It is silent on the converse question of whether macro determines micro. I will assume in what follows that it does not; that is, I'll assume that a given macro-state is *multiply realizable* at the micro-level. The relation of micro to macro is many-to-one. For example, an ideal chamber of gas can have a given temperature (its *mean* kinetic energy) by many different assignments of kinetic energy values to its constituent molecules.

The Laplacean picture of a deterministic universe is represented in Fig. 1. The MS principle says that a system's micro-state at one time fixes its macro-state at that time; A at t_1 makes it the case that X at t_1, and B at t_2 makes it the case that Y at t_2. This is the meaning of the vertical arrows in Fig. 1. The micro-state evolves by deterministic Newtonian rules (represented by the arrow from A to B), so the (complete) micro-state at one time fixes the micro-state at all later times. Since diachronic determination and synchronic supervenience are both necessitation relations, A at time t_1 insures that Y at t_2, by transitivity. If we use probabilities to describe whether macro-state Y will occur at time t_2, given the fact that the system was in macro-state X at t_1, we do so only because of our ignorance. The demon has no need of these probabilities. When the demon predicts the macro-state Y at t_2 from the micro-state A at t_1, it presumably first predicts the micro-state B at t_2 and then figures out that B suffices for the macro-state Y. Laplace's idea thus requires that the demon's knowledge extend beyond Newtonian matters; the demon also needs to be savvy about how Newtonian facts connect to facts described in the vocabularies of other sciences – for example, biology and psychology.

Given this Newtonian description of Laplace's demon, what would the corresponding picture be for a Laplacean who accepts indeterminism as a fact about the physical micro-level?[7] The situation is depicted in Fig. 2. If the micro-theory in question is indeterministic, the system's micro-state at t_1 confers probabilities on the different micro-states that might obtain at t_2. Some of these possible micro-states will be supervenience bases for the macro-property Y; others will not be. Suppose there are n disjoint micro-states $(B_1, B_2, \ldots, B_n)$ that are possible supervenience bases for the macro-state Y and that $\Pr(B_i \text{ at } t_2 | A \text{ at } t_1) = q_i$. We mere mortals, who are aware only of the macro-state X that obtains at time t_1, must predict whether Y

Fig. 2 Laplace's Demon in an Indeterministic Universe

	Time t_1	**Time t_2**
Macro-state	X —[p]→	Y
	$\uparrow$	$\updownarrow$
Micro-state	A —[q]→	(B_1 or B_2 or … or B_n)

will obtain at t_2 by computing the value of $p = \Pr(Y \text{ at } t_2 | X \text{ at } t_1)$. I will call this a *macro-probability* because the conditioning proposition describes the macro-state X that the system occupies at t_1. The demon, who can instantly see which micro-state obtains at t_1, will compute $\Pr(Y \text{ at } t_2|A \text{ at } t_1)$, which I term a *micro-probability*. Presumably the demon does this first by computing $q = q_1 + q_2 + \cdots + q_n = \Pr(B_1 \text{ or } B_2 \text{ or}, \ldots, \text{ or } B_n \text{ at } t_2|A \text{ at } t_1)$ and then taking into account the fact that the disjunction (B_1 or B_2 or, . . . , or B_n) is equivalent to Y. The point is that the value of $q = \Pr(Y \text{ at } t_2|A \text{ at } t_1)$ may differ from the value of $p = \Pr(Y \text{ at } t_2|X \text{ at } t_1)$. Although a Laplacean demon in an indeterministic universe will need to use probabilities to predict what will happen, it will use probabilities that may differ in value from the ones that we less-informed human beings are forced to employ. Were it not for our ignorance, we should do what the demon does, or so the Laplacean claims.[8]

The Laplacean position, then, does not depend on whether determinism is true. The first part of the position is a thesis about prediction:

(L_1) Suppose you know the system's macro-state (X) at t_1 and also know the system's micro-state (A) at t_1, and you want to predict whether the system will be in state Y at time t_2. If you know the values of both the macro-probability $\Pr(Y \text{ at } t_2|X \text{ at } t_1)$ and the micro-probability $\Pr(Y \text{ at } t_2|A \text{ at } t_1)$, and their values are different, then the micro-probability $\Pr(Y \text{ at } t_2|A \text{ at } t_1)$ is the one you should use to make your prediction.

Read contrapositively, this means that

> If you are entitled to use the macro-probability $\Pr(Y \text{ at } t_2|X \text{ at } t_1)$ in predicting whether Y will occur at t_2, this is because $\Pr(Y \text{ at } t_2|X \text{ at } t_1) = \Pr(Y \text{ at } t_2| A \text{ at } t_1)$ or you don't know the value of $\Pr(Y \text{ at } t_2|A \text{ at } t_1)$, or you don't know that A is the micro-state of the system at t_1.

This contrapositive brings out the fact that the Laplacean thinks that there are two possible justifications for using a macro-probability in making a prediction. One involves lack of knowledge; the other is the truth of a probabilistic equality.

The Laplacean principle (L_1) describes the probabilities you should use in making predictions, but does not connect that issue with the question of which probabilities are objective. This further element in the Laplacean position can be formulated as follows:

(L_2) If the only justification you have (or could have) for using the macro-probability Pr(Y at t_2|X at t_1) to predict whether Y will be true at t_2 is that you don't know the value of the micro-probability Pr(Y at t_2|A at t_1) or you don't know that A is the micro-state of the system at t_1, then the macro-probability Pr(Y at t_2|X at t_1) is not objective.

Whereas (L_1) says that there are two possible reasons for using a macro-probability to make a prediction, (L_2) says that one of those reasons (lack of knowledge) should lead you to attach a subjective interpretation to the macro-probability. Together, these two principles entail the Laplacean thesis that the only way the macro-probability Pr(Y at t_2| X at t_1) can be objective is for it to have the same value as the micro-probability Pr(Y at t_2|A at t_1). If these two probabilities have different values, the macro-probability should not be taken to describe an objective matter of fact.[9]

In what follows, I'll describe what I think is wrong with (L_2); (L_1), as I'll explain, follows from two principles that I'll assume without argument are correct. After criticizing (L_2), I'll try to answer the following two questions: Why should we think that a given macro-probability is objective? And where do objective macro-probabilities come from? My goal is to provide a non-Laplacean account of the epistemology and metaphysics of objective macro-probabilities.

The Principle of Total Evidence and Mereological Supervenience

What could motivate the two-part Laplacean position (L_1 and L_2)? It might seem to be an instance of the prejudice that Wilson (2004) calls *smallism* – the idea that it is better to conceptualize the world in terms of parts than in terms of wholes. In fact, the first conjunct in this two-part position (L_1) can be justified in terms of principles that many philosophers find compelling. L_1 follows from the *principle of total evidence* and the *principle of synchronic mereological supervenience*, both of which I'll assume are true for the rest of this paper. The principle of total evidence has nothing explicitly to say about micro and macro; rather, it bids you conditionalize on *all* of the information at your disposal when you try to figure out what the probability is of a future event Y. In particular, if you know both Φ and Ψ, and Φ entails Ψ, then using $Pr(Y|\Psi)$ will be a mistake if $Pr(Y|\Phi) \neq Pr(Y|\Psi)$. Applied to the problem at hand, the principle of total evidence says that if your micro-description A of the state of the system at time t_1 *entails* your macro-description X of the system at that same time, and the two descriptions confer different probabilities on the system's occupying state Y at time t_2, then you should use the former.[10,11] The (MS) principle completes the argument by affirming the antecedent; it asserts that A at t_1 entails X at t_1, if A is a *complete* description of the system's micro-state at t_1. Notice that this justification of (L_1) does not assert that *every* true micro-description is preferable to *every* true macro-description. If the micro-description is *incomplete*, it may or may not entail the macro-description in question.

Must (In)Determinism Percolate Up?

The thesis of synchronic mereological supervenience (MS) places a constraint on how determinism or indeterminism at one level is relevant to the same distinction at another:

$$\text{If Micro entails Macro and } \Pr(\text{Y at } t_2 | \text{ Macro at } t_1) = 1 \text{ (or 0),}$$
$$\text{then } \Pr(\text{Y at } t_2 \mid \text{ Micro at } t_1) = 1 \text{ (or 0).}$$

The reason this principle is true is that the probabilities 1 and 0 are *sticky*. If $\Pr(E|\Psi) = 1$ (or 0), then strengthening the conditioning proposition (by substituting Φ for Ψ, where Φ entails Ψ) cannot budge that value.[12] The principle holds regardless of whether the description of the micro-state is complete. Once again, the contrapositive is interesting:

(P) If Micro entails Macro and $\Pr(\text{Y at } t_2 | \text{ Micro at } t_1) \neq 1$ (or 0), then $\Pr(\text{Y at } t_2 | \text{ Macro at } t_1) \neq 1$ (or 0).

Proposition (P) is one way to express the idea that *indeterminism must percolate up*. This percolation principle is a consequence of the axioms of probability; it is not a consequence of those axioms that *determinism* must percolate up.

Notice that the percolation principle (P) has the micro- and the macro-descriptions probabilifying the same proposition, namely Y at t_2. If quantum mechanics says that your going to the movies tonight has a probability that is strictly between 0 and 1, then belief/desire psychology cannot assign that event a probability of 0 or 1, if your psychological state supervenes on your quantum mechanical state. Proposition (P) does not assert that if quantum mechanics assigns to *micro*-events probabilities that are strictly between 0 and 1, then psychology must assign intermediate probabilities to the *macro*-events it describes. That would be to make the following false claim (where, as before A and B_i are micro-properties and X and Y are macro):

$$\text{If } \Pr(B_i \text{ at } t_2 | \text{A at } t_1) \neq 1 \text{ (or 0)}, \quad \text{for each } i = 1, 2, \ldots, n,$$
$$\text{then } \Pr(\text{Y at } t_2 | \text{X at } t_1) \neq 1 \text{ (or 0).}$$

Even if A says that each B_i has an intermediate probability, it may still be true that A says that the disjunction (B_1 or B_2 or, ..., or B_n) has a probability of unity. This is how micro-indeterminism about the relationship of A to each B_i can be compatible with macro-determinism concerning the relation of X and Y. In *this* sense, micro-indeterminism need not percolate up.

If the axioms of probability are a priori, as I will assume, then so is (P). The truth of this percolation principle does not depend on anything empirical – for example, on the fact that radiation sometimes causes mutations that change the evolutionary trajectories of populations. Discussion in philosophy of biology of whether micro-indeterminism must percolate up into evolutionary processes has often focused on

	Time t_0	Time t_1	Time t_2
Macro-state		X	Y
Micro-state	Q	A	

Fig. 3 Two questions about percolation: (i) If A at t_1 entails X at t_1 and Pr(Y at t_2|A at t_1) is intermediate, must Pr(Y at t_2|X at t_1) also be intermediate? (ii) Can Q at t_0 affect Y at t_2 by affecting X at t_1? Question (i) can be answered a priori; (ii) cannot

empirical questions of this sort; see, for example, Brandon and Carson (1996), Glymour (2001), and Stamos (2001). To grasp the difference between these two questions about percolation, consider Fig. 3. My question concerns how the relationship of A at t_1 to Y at t_2 constrains the relationship of X at t_1 to Y at t_2, assuming that X supervenes on A. The other question focuses on is whether Q can affect Y by affecting X.

Although I have interpreted proposition (P) as saying that indeterminism must percolate up, the proposition also asserts that *determinism must filter down*. If we had a true deterministic macro-theory, that would entail that there must be a true deterministic micro-theory. It is interesting that philosophers usually think of micro-theories as constraining macro-theories, but not vice versa. Can this asymmetry be justified by pointing to the fact that quantum mechanics provides various no-hidden-variable proofs, whereas there are no deterministic theories in macro-sciences that we are prepared to say are true? This is a question that merits further exploration.

Does Macro Screen-Off Micro?

If a macro-probability is to pass the Laplacean test for objectivity, the macro-description of the system at t_1 must capture all of the information in the micro-description at t_1 that is relevant to predicting the system's macro-state at t_2. Where X is a macro-property at t_1 and A is a micro-property at that same time, we can express this idea by using Reichenbach's (1956) concept of screening-off:

$$\text{(Macro SO Micro)} \quad \Pr(Y \text{ at } t_2 | X \text{ at } t_1) = \Pr(Y \text{ at } t_2 | X \text{ at } t_1 \;\&\; A \text{ at } t_1).$$ [13]

If the micro-state at one time *entails* the macro-state at that time, as the principle of mereological supervenience (MS) asserts, then the micro-state at t_1 screens off the macro-state at t_1 from the macro-state at t_2:

$$\text{(Micro SO Macro)} \quad \Pr(Y \text{ at } t_2 | A \text{ at } t_1) = \Pr(Y \text{ at } t_2 | X \text{ at } t_1 \;\&\; A \text{ at } t_1).$$

		Macro-theory is	
		Deterministic	Indeterministic
Micro-Theory is	Deterministic	Always	Never
	Indeterministic	Not Defined	Depends on details

Fig. 4 When is (Macro SO Micro) true?

These two screening-off principles are compatible.[14] Together they entail that:

$$\Pr(\text{Y at } t_2 | \text{A at } t_1) = \Pr(\text{Y at } t_2 | \text{X at } t_1).$$

This last equality asserts that if you want to predict the system's state at t_2 based on information about its state at t_1, it doesn't matter whether you use the micro- or the macro-description. The two descriptions deliver the same probability.

To see when, if ever, (Macro SO Micro) is true, there are four cases to consider, which are shown in Fig. 4. The macro-theory, which relates X at t_1 to Y at t_2, will either be deterministic or indeterministic. I previously have discussed the micro-theory as relating A at t_1 to B at t_2. For purposes of the present discussion, I'm going to understand it as relating A at t_1 to Y at t_2. This theory is "micro" in the sense that it uses the micro-property A to predict whether Y will obtain. The micro-theory, like the macro-theory, will be either deterministic or indeterministic. I assume that A at t_1 suffices for X at t_1; the micro-property is a supervenience base for the macro-property.

Suppose the macro-theory is deterministic (the first column in Fig. 4). If so, the micro-theory must be deterministic as well, owing to the stickiness of 1's and 0's; the lower-left hand box in Fig. 4 is ruled out. When both theories are deterministic, (Macro SO Micro) is correct. As an example, consider Putnam's (1975) well-known example of the peg and the board. Putnam describes a board that contains two holes; one is round and is 1 in. in diameter; the other is square, and is a little more than 1 in. on each side. The peg he describes is square and is 1 in. on a side. The peg fits through the square hole, but not the round one. Why? Putnam contends that the correct explanation is given by the macro-dimensions just cited. He further claims that a micro-description of the configuration of the molecules in the peg and board is either not an explanation, or is a terrible explanation. Putnam concludes from this that reductionism is false – a macro-story explains something that no micro-story can explain (or explain as well). I prefer a more pluralistic view of explanation, according to which micro- and macro-stories are both explanatory (Sober 1999a); Jackson and Pettit (1992) do too. However, the present point is that, in Putnam's example, the macro-facts about the peg and board screen-off the micro-facts from the fact to be explained. This is because the peg will *necessarily* pass through one hole but not the other, given the macro-dimensions (and the initial condition that the peg is pushed in the right way).

	t_1		t_2		t_3
Macro	A	—[h]⟶	B	—[h]⟶	C
Actual Micro-supervenience bases	a	—[1]⟶	b	—[1]⟶	c
Possible Micro-supervenience bases	r(a)	—[0]⟶	r(b)	—[0]⟶	r(c)

Fig. 5 The (probabilistic) second law of thermodynamics says that macro-state A goes, with high probability, to B, and B, with high probability, to C. If A is realized by a, B by b, and C by c, this transition must occur, but if A is realized by r(a), B by r(b), and C by r(c), it cannot

What happens when the macro-theory is probabilistic (the second column in Fig. 4)? If the micro-theory is deterministic, the (Macro SO Micro) principle is false. Clearly, if the micro-probability Pr(Y at t_2|A at t_1) = 1 (or 0), and the macro-probability Pr(Y at t_2|X at t_1) has an intermediate value, then Pr(Y at t_2|X at t_1) $\neq$ Pr(Y at t_2|X at t_1 and A at t_1).[15] A situation of this type arises when the second law of thermodynamics is formulated probabilistically (saying that the entropy of a closed system at t_2 is, with high probability, no less than the entropy at t_1) and the underlying Newtonian micro-theory is taken to be deterministic. Figure 5 depicts an example described by Albert (2000, pp. 71–73), which he uses to explain ideas due to Zermelo and Loschmidt. A system moves from macro-state A at t_1 to B at t_2 and then to C at t_3, increasing in entropy at every step. A might be an isolated warm room that contains a block of ice, B the same room with a half-melted block of ice and a puddle, and C the room with no ice but a bigger puddle. Suppose that macro-state A happens to be realized by micro-state a, B by b, and C by c, and that the Newtonian laws of motion entail that a necessitates b and b necessitates c. Albert observes that the time-reverse of a, r(a), is a possible realizer of A (he uses the word "compatible"), r(b) a possible realizer of B, and r(c) of C, and that the laws of motion say that a state beginning in r(c) will necessarily move to r(b) and then to r(a). Given this, the macro-state B at t_2 does not screen off its micro-realization b from the macro-state C at t_3. Pr(C at t_3|B at t_2) is intermediate, but Pr(C at t_3|b at t_2) = 1 while Pr(C at t_3|r(b) at t_2) = 0.

The last case to consider is the one in which both the micro- and the macro-theories are indeterministic. In this case, I doubt that there can be a general argument to show that (Macro SO Micro) must *always* be false. However, with one type of exception, I know of no macro-descriptions that screen-off in the way that (Macro SO Micro) requires. The following two examples exemplify very general circumstances in which (Macro SO Micro) is false.

First, consider the macro-statement that Pr(lung cancer at t_2| smoking before t_1) = x, where "smoking" means smoking 10,000 cigarettes. Suppose there is a single carcinogenic micro-ingredient A in cigarette smoke and that different cigarettes contain different amounts of that micro-ingredient. This means that smoking 10,000 cigarettes entails inhaling somewhere between g and g+h grams of the carcinogenic ingredient. Given this, the macro-probability Pr(lung cancer at t_2| smoking before t_1) is a *weighted average* over the different probabilities of cancer that different levels of exposure to the micro-constituent A induce:

$$\begin{aligned}&\mathrm{Pr}(\text{lung cancer at } t_2 | \text{smoking before } t_1) = \\ &\quad \int_{n=g}^{n=(g+h)} \mathrm{Pr}(\text{lung cancer at } t_2 | \text{inhaling n grams of A before } t_1) \\ &\quad (\mathrm{Pr} \text{ inhaling n grams of A before } t_1 | \text{smoking before } t_1)(d\mathrm{n}).\end{aligned}$$

If the risk of cancer is an increasing function of the number of grams of A that you inhale, then the macro-probability will have a different value from all, or all but one, of the micro-probabilities:

$$\begin{aligned}&\mathrm{Pr}(\text{lung cancer at } t_2 | \text{smoking before } t_1) \neq \\ &\quad \mathrm{Pr}(\text{lung cancer at } t_2 | \text{smoking before } t_1 \,\& \\ &\quad \text{inhaling n grams of A before } t_1) \text{ for all, or all but one, value of n.}\end{aligned}$$

The point is really very simple: if all the children in a classroom have different heights, at most one of them will have a height that is identical with the average height.[16,17]

The one kind of case I know of in which (Macro SO Micro) is true where the macro-theory is not deterministic involves a macro-description of the system at t_1 that is *defined* so as to confer a certain probability on a macro-state at t_2. As mentioned earlier, the viability component of an organism's fitness is defined as its probability of surviving from egg (at t_1) to adult (at t_2). This means that

$$\begin{aligned}&\mathrm{Pr}(\text{O is alive at } t_2 | \text{O is alive at } t_1 \text{ and has a viability} = \text{x}) = \\ &\quad \mathrm{Pr}(\text{O is alive at } t_2 | \text{O is alive at } t_1 \text{ and has a viability} = \text{x} \,\&\, \text{O} \\ &\quad \text{at } t_1 \text{ has genotype G}) = \text{x}.\end{aligned}$$ [18,19]

I take it that the genotypic description is a "micro-description," as compared to the fitness description, which is more macro, since the latter attaches to the whole organism without mentioning its parts. Perhaps a radioactive atom's *half-life* provides a similar example; it is defined so that the screening-off relation holds.

Natural selection might seem to provide the perfect setting for (Macro SO Micro) to be true. It is often claimed that natural selection "cares" only about an organism's phenotype, and not about its genotype; in this vein, Mayr (1963, p. 184), Gould (1980, p. 90), and Brandon (1990) have emphasized the idea that natural selection acts "directly" on phenotypes, and only indirectly on genotypes. Perhaps their point should be formulated in the way that (Macro SO Micro) suggests:

$$\begin{aligned}&\mathrm{Pr}(\text{O is alive at } t_2 | \text{O has phenotype P at } t_1) = \mathrm{Pr}(\text{O is alive at } t_2 | \text{O} \\ &\quad \text{has phenotype P at } t_1 \text{ and genotype G at } t_1).\end{aligned}$$

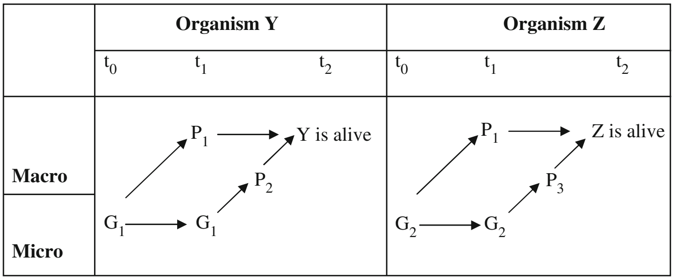


Fig. 6 A Failure of Screening-Off. Phenotype at t_1 does not screen off genotype at t_1 from being alive at t_2

In fact, the temporal gap between t_1 and t_2 provides room for this relationship to be falsified, as Fig. 6 illustrates. Consider two individuals (Y and Z) who have different genotypes (G_1 and G_2); they have the same total phenotype (P_1) at t_1 but different chances of surviving until t_2. The reason this is possible is that Y's genotype causes it to develop phenotype P_2 at a time after t_1 but before t_2, while Z's genotype causes it to develop phenotype $P_3 (P_2 \neq P_3)$ at that intermediate time. An organism's *entire suite* of phenotypic traits up until t_2 affects its chance of surviving until t_2. Although this entire suite may screen-off genotype from survival (this is a better way to put the Mayr/Gould/Brandon point), the phenotypic traits that an organism has *at a single time* during its development often will not.

The arrows in Fig. 6 all represent causal relations. I assume that an organism's genome is completely stable throughout its lifetime, though this isn't essential for my point; allowing for mutation would mean that an organism's genome at one time exerts a probabilistic (not a deterministic) influence on its genome later. I also assume that cause must precede effect; this is why no arrow connects the genotype at t_1 to the phenotype at t_1. However, the argument against (Macro SO Micro) does not depend on ruling out simultaneous causation. It also does not matter whether genotype at t_0 causally determines phenotype at t_1; this means that even if genotype at t_1 provided a supervenience base for phenotype at t_1, there still would be a counterexample to (Macro SO Micro).

Figure 6 depicts a general circumstance in which (Macro SO Micro) fails. The reason the macro-state at t_1 does not screen-off the micro-state at t_1 from the macro-state at t_2 is that there are *two causal pathways* from the micro-state at t_0 to the macro-state at t_2. The macro-state at t_1 occurs on just one of them.[20] A predictor, demonic or human, who wants to say whether an organism will be alive at t_2 by using information about the organism's state at t_1 may do better by using the organism's genotype at t_1, rather than the organism's phenotype at t_1, as the basis for the prediction.[21]

A similar argument undermines the claim that phenotype screens-off genotype from reproductive success when an organism's reproductive success is defined as its expected number of *viable* offspring:

$$\text{E(number of viable offspring that O has} | \text{O has phenotype P)} =$$
$$\text{E(number of viable offspring that O has} | \text{O has phenotype P \& O has genotype G).}$$

For a counterexample, consider a dominant gene A; AA and Aa individuals both have phenotype P, which differs from the phenotype Q that aa individuals possess. Suppose individuals with P have higher viability than individuals with Q. If so, AA individuals will have a greater expectation of viable offspring than do Aa individuals, even though they are phenotypically identical. This is because Aa individuals sometimes have aa offspring, but AA individuals never do (Sober 1992). The two-pathway pattern is present here as well. An individual's genotype influences its own phenotype as well as the genotype its offspring have. This is why screening-off fails.[22]

It is important to bear in mind that my pessimistic evaluation of (Macro SO Micro) is predicated on the assumption that the principle of mereological supervenience (MS) is true. If this assumption is dropped, all four cells in Fig. 4 must be reconsidered; for example, indeterministic micro-theories can be consistent with deterministic macro-theories, since the stickiness argument no longer applies. Forster and Kryukov (2003) point out that investigations of the relation of micro- to macro-theories in physics often conceptualize macro-states as probabilistic expectations over possible micro-states; this means that the system's (actual) macro-state at a time is not determined by its (actual) micro-state at that time in any obvious way. Quantum mechanics has forced philosophers to take seriously the possibility that the diachronic thesis of determinism may be false. Perhaps the synchronic determination thesis (MS) should be re-evaluated as well (Crane and Mellor 1990; Sober 1999b).

Micro- and Macro-Causation

In the previous section I argued that the micro-state of a system at t_1 is often correlated with its macro-state at t_2, even after you control for the system's macro-state at t_1. This is what it means for the (Macro SO Micro) principle to fail. I now want to argue that this relationship between the micro-state at t_1 and the macro-state at t_2 often isn't a *mere* correlation; the micro-state at t_1 often is a *cause* of the macro-state at t_2.

To defend this claim, I want to exploit the suggestive ideas about causation presented by Woodward (2003), who draws on the frameworks developed by Spirtes et al. (2000) and Pearl (2000). When two events X and Y are correlated, how are we to discriminate among the following three possibilities: (i) X causes Y; (ii) Y causes X; and (iii) X and Y are joint effects of a common cause C?[23] The intuitive idea is that if X causes Y, then intervening on X will be associated with a change in Y; this won't be true if Y causes X, or if X and Y are effects of a common cause C. To make this suggestion precise, one has to define the concept of an intervention very carefully, which Woodward does. An intervention on X with respect to Y causes the state of X to take on a particular value; it therefore cancels the other causal influences that would otherwise impinge on X. In addition, an intervention must be delicate, not ham-fisted; if X and Y are joint effects of a common cause C, then an intervention on X with respect to Y must fix the state of X without simultaneously modifying the state of C. Woodward points out that "intervention" is a causal concept, so the account of causation he gives is not reductive. However, the account is not circular, in that the intervention criterion for when X causes Y does not require that you already know whether X causes Y. Even so, it is a consequence of Woodward's theory that you must have lots of *other* causal knowledge if you are to figure out whether X causes Y.

In the first example discussed in this section smoking and lung cancer are the macro-variables, and the number of grams inhaled of the carcinogenic micro-ingredient A is the micro-variable. It seems clear here that intervening on the micro-variable while holding fixed the fact that someone smokes will change the person's probability of getting lung cancer. This tells you that the micro-state at t_1 causally contributes to the macro-state at t_2. The same pattern obtains in the example depicted in Fig. 5 in which phenotype at t_1 and being alive at t_2 are the two macro-variables, and genotype at t_1 is the micro-variable. If we change someone's genotype (from G_1 to G_2, or vice versa), while leaving the phenotype unchanged, the chance of surviving until t_2 will change.

These remarks in favor of the micro-state at t_1's being causally efficacious do not rule out the possibility that the macro-state at t_1 is also causally efficacious (Sober 1999b). An intervention that shifted someone from 10,000 cigarettes smoked to, say, 25, would be associated with a reduction in that person's chance of getting lung cancer. Of course, this drastic reduction in number of cigarettes smoked will entail a drastic reduction in how many A particles are inhaled. And it may also be true that smoking causes cancer only because cigarette smoke contains A particles. But none of this should be taken to refute the claim that smoking causes cancer. There is no conflict between the claim that smoking causes cancer *and* the claim that inhaling A particles causes cancer.[24,25]

The Principle of Total Evidence and Explanation

Suppose we use the macro-probability $\Pr(Y \text{ at } t_2|X \text{ at } t_1)$ to predict whether Y at t_2 because we know that X is the system's macro-state at t_1 and we don't know the value of the micro-probability $\Pr(Y \text{ at } t_2|A \text{ at } t_1)$ and also don't know the system's micro-state at t_1. The Laplacean concludes from this that we should give a subjective interpretation to the macro-probability $\Pr(Y \text{ at } t_2|X \text{ at } t_1)$. We now need to evaluate this piece of advice, embodied in the principle L_2. After all, the Laplacean principle L_1 has to do with which probabilities we should use *to make predictions*. If we use probabilities for other purposes – for example, to construct explanations – perhaps this different venue can provide a reason for thinking that macro-probabilities are objective (Sober 1984).

Does our interest in constructing *good explanations* justify our citing X at t_1 as an explanation of Y at t_2, rather than citing A at t_1 instead, when $\Pr(Y \text{ at } t_2|X \text{ at } t_1) \neq \Pr(Y \text{ at } t_2|A \text{ at } t_1)$? To endorse this suggestion, we need not agree with Putnam's claim that the micro-details are explanatorily irrelevant. As argued in the previous section, it is often true that both the micro- and the macro-properties of a system at a given time are causally efficacious. If explanation means causal explanation, then causal explanations can be constructed in terms of micro-properties and also in terms of macro-properties. Is there a sense in which the macro-explanations are objectively better?

Putnam, following Garfinkel (1981), holds that a good explanation will be general, and that more general explanations are objectively better than less general ones. If the micro-property A entails the macro-property X, but not conversely, then X will be true of at least as many systems as A is. This is what Garfinkel and Putnam mean by generality. With respect to Putnam's peg and board, there will be more systems that have the macro-dimensions he describes than have the exact molecular configuration that the micro-description specifies.

My reply is that generality is just one virtue an explanation might have, and that we want different explanations to exhibit different virtues, or different mixes of them. Sometimes we want greater generality; at others we want more causal detail. For example, suppose we are evaluating two explanations of an event; one candidate cites just one of its causes while the other cites two; the second of these will be less general than the first, but will be more detailed. The *desiderata* of breadth and depth of explanation conflict (Jackson and Pettit 1992; Sober 1999a) and there is no objective criterion concerning which matters more or that determines what the optimal trade-off between them is.

Even when we focus just on generality, it isn't automatic that macro-explanations trump micro-explanations. First, the Garfinkel–Putnam definition of generality entails that the macro-description of a system at a given will be more general than a micro-description of the system at that time, *when the latter entails the former*. The principle of mereological supervenience (MS) assures us that the two descriptions will be related in this way when the micro-description is *complete*. However, this leaves unsettled which description is more general when the micro-description of the system's state is *not* complete. Garfinkel and Putnam have not shown that a macro-explanation of E is more general than *any* micro-explanation of E, but only that the macro-explanation is more general than a *complete* micro-explanation. Their argument against reductionism therefore fails, even if we accept their definition of generality. The second problem has to do with Garfinkel and Putnam's decision about how generality should be defined. Consider the causal relationships depicted in Fig. 7. The distal cause D causes each of $P_1, P_2, \ldots, P_n$, and each of these proximate causes has its own effect $E_1, E_2, \ldots, E_n$. Let's suppose that D suffices, but is not necessary, for P_2. The Garfinkel–Putnam definition of generality will then say that P_2 is a more general explanation of E_2 than D is. Their definition of generality focuses on the *explanandum* E_2 and asks how many systems that have this target property have D and how many have P_2. However, there is another definition of generality, one which focuses on how many phenomena a given *explanans* explains. D explains all of the P_i, and all of the E_i as well, whereas P_2 explains only E_2. In this sense, D is more general than P_2 (Tsai 2004). There is no conflict here, of course; P_2 is a more general explanation than D is *of a single explanandum* (namely E_2), but D

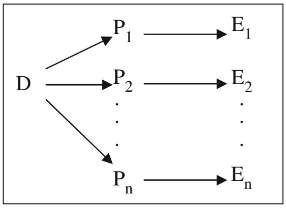


Fig. 7 P_2 may be a more general explanation of E_2 than D is, but D explains more phenomena than P_2 does

applies to more *explananda* than P_2 does. The problem is that the Garfinkel–Putnam argument depends on using one definition of generality and ignoring the other.[26]

The Laplacean principle L_1 is correct as a claim about prediction, assuming as I do that the principle of mereological supervenience (MS) and the principle of total evidence are both correct. However, the objectivity of macro-probabilities is not to be secured by pointing out that one of our goals in constructing explanations is that they be general. Sometimes we want our explanations to be more general; at other times we want them to be more detailed. But in both cases, we want the facts we cite to be facts – we want them to be objectively correct. To echo the *Euthyphro*, our citing a macro-probability when we give an explanation is not what makes that probability objective; rather, its objectivity is a requirement we impose on the items we choose to include in our explanations. So even if generality of explanation were a categorical imperative, that would not show that macro-probabilities are objective. And, in any event, generality is just one explanatory virtue among several.

The Principle of Total Evidence, Objectivity, and the Smart Martian Problem

The Laplacean argument against the objectivity of macro-probabilities resembles an argument that Nozick constructed (described in Dennett 1987) concerning the reality of beliefs and desires. Nozick pointed out that smart Martians will be able to predict our behavior without needing to attribute beliefs and desires to us. They can grasp at a glance the properties of the elementary particles that make up our bodies and use that description as an input to a dynamical physical model to predict how our bodies will comport themselves in the future. Nozick took this to raise the question of why we should think that people really do have beliefs and desires, a question that Dennett (1987) answered, following the Garfkinkel–Putnam line, by citing our penchant for constructing general explanations. The assumption behind Nozick's puzzle is that the only reason there could be for thinking that individuals have beliefs and desires is that we need to postulate those states to predict behavior.[27] We mere mortals need to attribute beliefs and desires if we wish to predict a person's behavior, but smart Martians do not. This seems to rob the existence of beliefs and desires of their objectivity.

The main thing wrong with this argument is its ad hominem quality. If you want to know whether something exists, you should ask to see the relevant evidence.

Whether you or anyone else *needs* to believe that the thing exists for purposes of prediction (or explanation) is not separately relevant. Nozick's demon won't need to say that bowling balls exist, but that hardly shows that there are none (Sober 1999b). The same holds for the existence of beliefs and desires, and also for objective macro-probabilities.[28]

The comparison that Nozick invites us to make is between two hypotheses, one ascribing a set of beliefs and desires to someone, the other ascribing some complex micro-physical state to that person. According to Nozick's story, the Martian knows that the second of these hypotheses is true, and so he has no need to consider the first. This fact is supposed to raise the question of why we should think that people really have beliefs and desires. Why shouldn't we treat this posit instrumentally, as a fiction that we find useful? However, if the question is whether a person really has some set of beliefs and desires, the appropriate alternative hypothesis to consider is not that they occupy some complex micro-physical state. These are not competing hypotheses – both could well be true. To treat them as competitors would be like wondering whether someone has smoked cigarettes, and then taking the alternative to be that he has inhaled carcinogenic A particles.

If the Laplacean argument against the reality of macro-probabilities were sound, it would be possible to strengthen it. Consider a hypothetical being that has perfect precognition. Unlike Laplace's demon, it doesn't need to observe the present state of the universe and then compute its future; this being knows the whole history – past present, and future – directly. This super-demon would not need to use any dynamical law – micro or macro, deterministic or indeterministic – to predict the future. But surely this does not show that dynamical laws (e.g., those of quantum mechanics) are never objectively true – that they are just useful fictions. Here again, what a hypothetical demon needs does not settle what is objectively true.

The mistaken idea that the Laplacean criterion (L_2) is a good test for whether a probability is objective may seem to receive support from the equally mistaken assumption that a proposition has just one true probability. This second mistake leads to the pseudo-problem of trying to figure out what that one true probability is. The grip of this misconception can be broken by taking conditional probability, rather than unconditional probability, as one's fundamental concept (Hajek 2003).[29] The proposition that Y occurs at t_2 has a different probability, depending on which conditioning proposition is chosen. To ask whether $Pr(Y \text{ at } t_2 | X \text{ at } t_1)$ or $Pr(Y \text{ at } t_2 | A \text{ at } t_1)$ is the true probability of Y at t_2 is like asking what the true distance is to Madison, the distance from Baltimore to Madison or the distance from Boston to Madison.[30]

The Epistemology and Metaphysics of Objective Macro-probabilities

If the needs of demons are not relevant to finding out whether a probability is objective, where should we look? We may begin by asking what it means to say that the macro-probability $Pr(Y \text{ at } t_2 | X \text{ at } t_1)$ or the micro-probability $Pr(Y \text{ at } t_2 | A \text{ at } t_1)$

describes objective features of the world. Under the heading of "objective interpretations of probability," there is a now-familiar list – actual relative frequency, hypothetical relative frequency, and propensity, each with variations on those broad themes (Eells 1981). If *any* of these interpretations is correct for *any* statement that assigns a value to a macro-probability, the objectivity of that statement is vouchsafed. For example, if Pr(this coin lands heads | this coin is tossed) is taken to describe the actual frequency of heads in a run of tosses, there is nothing subjective about a statement that assigns this probability a value. It is perfectly possible for the statement to be objectively true.[31]

Unfortunately, this does not suffice to show that macro-probabilities are objective, since none of these interpretations is adequate as a general account of objective probability. The objections I have in mind are familiar. With respect to the actual frequency interpretation, the fact is that we often conceptualize probabilities in such a way that they can have values that differ from actual frequencies; for example, a fair coin can be tossed an odd number of times and then destroyed. The other objective interpretations fare no better. If propensities are causal tendencies – that is, if Pr(Y at t_2| X at t_1) represents the causal tendency of X at t_1 to produce Y at t_2 – then the propensity interpretation cannot make sense of the "backwards probabilities" mentioned at the start of this paper that have the form Pr(X at t_1| Y at t_2), at least not if cause must precede effect (this objection is due to Paul Humphreys; see Salmon 1984, p. 205).[32] On the other hand, if the concept of propensity is stripped of this causal meaning, it isn't clear how the propensity interpretation helps clarify the concept of objective probability. As for the hypothetical relative frequency interpretation, it overstates the relation of probability to what would happen in the infinite long run. It is possible for a fair coin to land heads each time in an infinite run of tosses (though this, like all the other exact sequences that can occur, has a probability of 0). The coin's probability of landing heads is *probabilistically* (not *deductively*) linked to what would happen in the long run, finite or infinite (Skyrms 1980); the hypothetical relative frequency interpretation therefore does not provide a reductive definition of objective probability.

In view of the failures of these interpretations, my preference is to adopt a *no-theory theory of probability*, which asserts that objective probability is not reducible to anything else. Frequencies provide evidence about the values of probabilities, and probabilities make (probabilistic) predictions about frequencies, but probabilities don't reduce to frequencies (Levi and Morgenbesser 1964; Levi 1967; Sober 1993b, 2003b). Instead, we should view objective probabilities as theoretical quantities. With the demise of logical positivism, philosophers abandoned the idea that theoretical magnitudes such as mass and charge can be reduced to observational concepts that are theory-neutral. We should take the same view of objective probabilities.

If we reject the need for a reductive interpretation of objective probability, what does it mean to say that a probability is objective? Taking our lead from other theoretical concepts, we can ask what it means to say that mass is an objective property. The idea here is that mass is a mind-independent property; what mass an object has does not depend on anyone's beliefs or state of mind.[33] The type of independence involved here is conceptual, not causal – it is not ruled out that an object have the mass it does because of someone's beliefs and desires. The next question we need to ask is epistemological – what justifies us in thinking that mass is an objective property? If different measurement procedures, independently put to work by different individuals, all lead to the same estimate of an object's mass, that is evidence that mass is an objective property. The matching of the estimates is evidence that they trace back to a common cause that is "in" the object; since the estimates do not vary, differences among the procedures used, and among the psychological states of the investigators, evidently made no difference in the results obtained. Of course, this sort of convergence does not *prove* that there is an objective quantity that the observers are separately measuring.[34] It is possible that the observers have some psychological property in common, and it is this subjective commonality that causes their estimates to agree, there being no mind-independent reality that they are bumping up against at all.[35]

A perfect matching of the estimates that different investigators obtain is not necessary for the common-cause argument I am describing to go forward. Suppose the different estimates differ, but only a little. That too would support an argument for there being an objective quantity that the different investigators are measuring. Put formally, this amounts to endorsing a model that says that there exists a true value of the mass of the object in question, and that the observers obtain their estimates by processes that are characterized by a set of error probabilities. Endorsement of this model should be understood as a comparative claim, not an absolute claim – we are judging that this model is *better* than one or more alternatives. The alternative of interest here is a model that says that each investigator is measuring a separate property of his or her state of mind. This more complex model can be made to fit the data, but its greater complexity counts against it. In this way, the question of whether mass is objective can be turned into a problem of model selection.[36]

How can we use the example of mass to guide our thinking about the objectivity of macro-probabilities? To begin with, we must take account of a difference between mass and probability.

Mass is a non-relational, intrinsic property of an object, but, as mentioned earlier, I want to regard conditional probability, not unconditional probability, as the fundamental notion, and conditional probability is a relation between pairs of propositions. The case for the objectivity of macro-probabilities is not defeated by the fact that you and I will give different probabilities for a coin's landing heads if we conditionalize on different information about the system's initial conditions. It is obvious that $Pr(Y|X)$ and $Pr(Y|A)$ can have different values. Our question is whether $Pr(Y|X)$ and $Pr(Y|A)$ can both be objective quantities.

It may seem that focusing on conditional probabilities does not change matters. If you have background knowledge U and I have background knowledge M, then you will evaluate $Pr(Y|X)$ by computing $Pr(Y|X\&U)$, while I'll evaluate $Pr(Y|X)$ by computing $Pr(Y|X\&M)$. If $Pr(Y|X\&U) \neq Pr(Y|X\&M)$, you and I will assign different values to $Pr(Y|X)$. This suggests that $Pr(Y|X)$ does not represent an objective relation that connects the propositions Y and X; rather, there is a third *relatum* that is not explicitly mentioned, and this is the set of beliefs that some agent or other possesses. However, if this is the right take on Pr(Y|X), what should we say about Pr(Y|X&U) and Pr(Y|X&M)? Do *they* have unique values, or are their values again relative to the background beliefs of some agent? If the values aren't unique, how did you and I manage to assign values to Pr(Y|X) by consulting our background knowledge U and M and then deciding what the values are for Pr(Y|X&U) and Pr(Y|X&M)? In fact, what gets added to X and Y when U and M are taken into account is not anyone's psychological state, but certain further propositions. It isn't *your believing* U that matters, but simply the proposition U that you believe; this is what gets included as one of the conditioning propositions in Pr(Y|X&U). These considerations suggest the following thesis: *some* conditional probabilities have unique values, which are not relative to anyone's background beliefs. This claim is of course compatible with the concession that *many* conditional probabilities do not have this status.[37] Statisticians have a term for probabilities of the first sort – these are the probabilities that figure in "simple" (as opposed to "composite") statistical hypotheses.

To develop my argument for the reality of (some) macro-probabilities, I want to consider a Newtonian model of coin tossing due to Keller (1986) and Diaconis (1998). The initial conditions for a toss determine whether the coin will land heads or tails. The reason a coin exhibits some mixture of heads and tails in a series of tosses is that the initial conditions vary from toss to toss. To simplify matters, we assume that there is no air resistance, that the coin spins around a line through its plane, and that the coin lands without bouncing (perhaps in sand). The relevant initial conditions are then fixed by specifying the values of V (the upward velocity of the toss) and ω (the angular velocity, given in revolutions per second). If V is very low, the coin doesn't go up much; if ω is very low, the coin, as Diaconis puts it, "rises like a pizza without turning over." Depending on the values of V and ω, the coin will turn 0, 1, 2, 3, . . . times before it lands. Suppose the coin we are considering starts each tossing session by being heads up in the tosser's hand. Then, if the coin turns over 0 or an even number of times, it lands heads, and if it turns over an odd number of times, it lands tails. These different possibilities correspond to the regions of parameter space depicted in Fig. 8. Starting at the origin and moving Northeast, the different stripes correspond to 0 turns, 1 turn, 2 turns, etc. For the purposes of investigating how macro-probabilities can be objective in a deterministic universe, I'm going to assume that the Keller/Diaconis model is true.[38] Laplace's demon, since it knows the exact values of V and ω that characterize a given toss, will be able to predict whether the coin will land heads or tails without needing to use the concept of probability. We who are less well informed about the toss's values for V and ω need to use the language of probability to make our prediction. This much is uncontroversial. The question is how a macro-probability, which does not conditionalize on point values for V and ω, can be objective.[39]

I want to argue, not just that the objectivity of macro-probabilities is *compatible* with Keller and Diaconis's deterministic model, but that this deterministic micro-model helps *justify* the claim that the macro-probabilities are objective. Consider the followingrepresentation of the relation between macro-and micro-dynamical

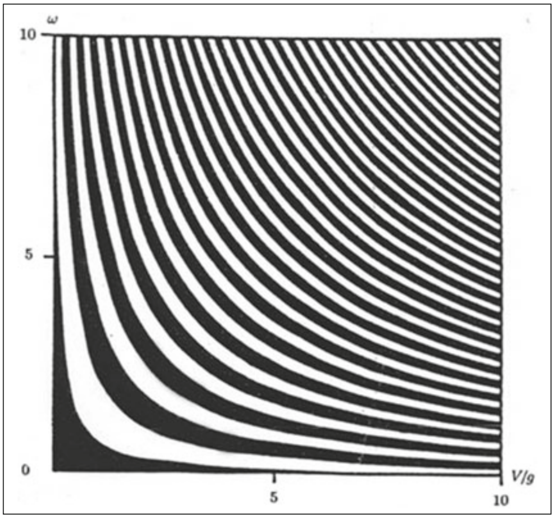


Fig. 8 Newtonian coin tossing (From Diaconis 1998)

theories. Suppose there are n micro-states ($A_1, A_2, \ldots, A_n$) the system might be in at t_1. As before, X and Y are macro-properties.

$$\text{(D)} \quad \Pr(\text{Y at } t_2 | \text{X at } t_1) = \sum_i \Pr(\text{Y at } t_2 | A_i \text{ at } t_1) \Pr(A_i \text{ at } t_1 | \text{X at } t_1).$$

As noted earlier, the macro-probability $\Pr(\text{Y at } t_2|\text{X at } t_1)$ is a weighted average over various micro-probabilities of the form $\Pr(\text{Y at } t_2|A_i \text{ at } t_1)$. The Laplacean grants that the first term on the right-hand side, $\Pr(\text{Y at } t_2|A_i \text{ at } t_1)$, is objective. So the question of whether the macro-probability $\Pr(\text{Y at } t_2|\text{X at } t_1)$ is objective reduces to a question about the second term on the right – is the distribution of $\Pr(A_i \text{ at } t_1|\text{X at } t_1)$ over $i = 1, 2, \ldots, n$ objective? That is, *the diachronic macro-probability $Pr(Y \text{ at } t_2|X \text{ at } t_1)$ is objective if a set of synchronic probabilities is too*.[40] This point holds regardless of whether the micro-theory is deterministic – i.e., independently of whether all the probabilities of the form $\Pr(\text{Y at } t_2|A_i \text{ at } t_1)$ have extreme values. Those with Laplacean sympathies may think that macro-probabilities can't be objective if the dynamic micro-laws are deterministic. But examining the decomposition (D) shows that this reaction involves looking in the wrong place. Even if the micro-level dynamics are deterministic, what matters is the *initial conditions* – the question is whether the distribution $\Pr(A_i \text{ at } t_1|\text{X at } t_1)$ is itself objective.

If we apply the decomposition (D) to the case of coin tossing, and take account of the fact that V and ω are continuous quantities, we obtain the following double integral:

$$\Pr(\text{heads at } t_2|\text{tossed at } t_1) = \iint_{a,b} \Pr(\text{heads at } t_2|\text{V} = \text{a and } \omega = \text{b at } t_1)$$
$$\Pr(\text{V} = \text{a and } \omega = \text{b at } t_1|\text{tossed at } t_1)(da)(db).$$

The question is whether the second product term on the right – the distribution of initial conditions – is objective. If it is, so is the macro-probability on the left. The Keller/Diaconis model helps simplify our question. Figure 8 shows that all we need to worry about is

$$\begin{aligned}\Pr(\text{heads at } t_2|\text{tossed at } t_1) = {} & \Pr(\text{heads at } t_2|\text{V and } \omega \text{ are in a black region at } t_1)\\ & \Pr(\text{V and } \omega \text{ are in a black region at } t_1|\text{tossed at } t_1),\end{aligned}$$

which simplifies to

$$\Pr(\text{heads at } t_2|\text{tossed at } t_1) = \Pr(\text{V and } \omega \text{ are in a black region at } t_1|\text{tossed at } t_1).$$

Let me emphasize once more that the diachronic macro-probability on the left is objective if the synchronic probability on the right is.

Why should we think that Pr(V and ω are in a black region at t_1|tossed at t_1) is an objective quantity? Suppose that a large number of observers each toss the coin one or more times. The number of tosses may vary from observer to observer and the observers may obtain somewhat different frequencies of heads in their different runs. Suppose that experimenters who toss the coin 1,000 times or more obtain frequencies of heads that are tightly clustered around 51%, while those who toss the coin a much smaller number of times obtain frequencies that are more widely dispersed. The similarity of the actual frequencies obtained by different observers should be explained by postulating a common cause. There is something about the process generating initial conditions on a toss that leads 51% of the tosses to be located in the black region of parameter space. This commonality is captured by the claim that Pr(V and ω are in a black region at t_1| tossed at t_1) = 0.51. This probability claim applies not only to the runs of tosses that have been performed to date; it additionally makes a prediction about future runs of tosses, a prediction in which we are entitled to have some confidence.[41]

The common cause argument just presented is epistemic – the point is to show that the kind of evidence we have concerning the objectivity of mass as a property also helps establish the objectivity of synchronic probabilities of the form Pr(A at t_1|X at t_1). This point about evidence, however, does not answer a more metaphysical question. Where do macro-probabilities come from? What account can be given of how they arise?

The idea of common causes provides an answer to this metaphysical question as well, one that is depicted in Fig. 9. As before, let X be a macro-property at time t_1 and A a micro-property at that same time, but now consider C, which is a micro-variable at the earlier time t_0. C is a common cause of X and A, and C comes in n states (C_1, C_2, ..., C_n). We are interested in the question of how the macro-probability Pr(A at t_1|X at t_1) arises. The answer[42] is that it arises from the common cause C:

$$\Pr(A|X) = \Pr(A\&X)/\Pr(X) = \sum_i \Pr(A\&X|C_i)\Pr(C_i) \Big/ \sum_i \Pr(X|C_i)\Pr(C_i).^{43}$$

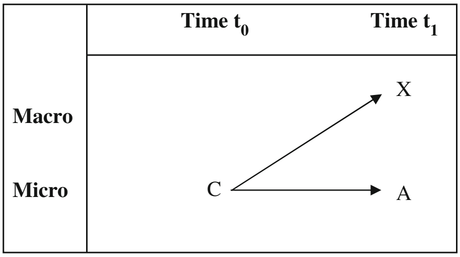


Fig. 9 How a synchronic macro-probability can arise from a micro common cause

Notice that all the probabilities on the right of the second equality sign are conditional micro-probabilities or unconditional probabilities of the micro-properties C_i. A Laplacean who believes in the reality of micro-probabilities can see from this how synchronic macro-probabilities of the form $\Pr(A \text{ at } t_1|X \text{ at } t_1)$ arise. And once these are in place, the reality of diachronic macro-probabilities of the form $\Pr(Y \text{ at } t_2|X \text{ at } t_1)$ is vouchsafed. We thus have in hand both a metaphysical and an epistemological answer to our question about the reality of macro-probabilities.

Concluding Comments

The Laplacean position (L_1 and L_2) allows that there is one circumstance in which the macro-probability $\Pr(Y \text{ at } t_2|X \text{ at } t_1)$ represents an objective matter of fact. This occurs when the macro-probability $\Pr(Y \text{ at } t_2|X \text{ at } t_1)$ and the micro-probability $\Pr(Y \text{ at } t_2|A \text{ at } t_1)$ have the same value. If A at t_1 suffices for X at t_1, in accordance with the principle of mereological supervenience, this equality holds precisely when (Macro SO Micro) – that is, when $\Pr(Y \text{ at } t_2|X \text{ at } t_1) = \Pr(Y \text{ at } t_2|X \text{ at } t_1 \text{ and } A \text{ at } t_1)$. I have argued that this equality will rarely if ever be true. If we had true deterministic macro-theories, (Macro SO Micro) would automatically be correct. But we do not. And when both the macro- and the micro-theories are probabilistic, I know of no cases in which (Macro SO Micro) is correct; what is more, there seem to be very general reasons why this principle should fail. Facts about the micro-level seem constantly to provide predictively relevant information – information that is relevant above and beyond that provided by more coarse-grained macro-facts.

If (Macro SO Micro) were true, that principle would provide macro-probabilities and the theories in which they figure with a degree of autonomy. Predicting whether Y at t_2 would require just the macro-information that X at t_1; adding micro-details about what is true at t_1 would not be relevant. Anti-reductionists who yearn for this type of autonomy need to face up to the fact that it is not to be had. The autonomy of macro-level theories, if it exists, must be found elsewhere.

Laplaceans reason that if (Macro SO Micro) is false, then the macro-probability $\Pr(Y \text{ at } t_2|X \text{ at } t_1)$ fails to describe anything objective. If we want to resist this conclusion, what are our options? One is to appeal to the usefulness of macro-probabilities in constructing explanations. I have offered two reasons for rejecting this suggestion. First, it is unclear why explanations constructed by using macro-probabilities are *objectively* better than explanations constructed on the basis of micro-probabilities. And second, it is a mistake to think that macro-probabilities are objective *because* they figure in our explanations. Rather, the situation is the reverse – if we want our explanations to be objectively correct, we can cite macro-probabilities only if we have some assurance that they are objective.

The line I prefer to take against the Laplacean focuses on the decomposition (D). If we assume, as the Laplacean does, that the micro-probabilities $\Pr(\mathrm{Y\ at\ t_2}|\mathrm{A_i\ at\ t_1})$ $(i = 1, 2, \ldots, n)$ are objective, it turns out that the macro-probability $\Pr(\mathrm{Y\ at\ t_2}|\mathrm{X\ at\ t_1})$ is objective precisely when the synchronic probabilities $\Pr(\mathrm{A_i\ at\ t_1}|\mathrm{X\ at\ t_1})(i = 1, 2, \ldots, n)$ are too. Even if the micro-dynamical laws are deterministic, it is irrelevant to harp on that fact. The crux of the matter concerns the synchronic distribution of initial conditions – in Keller and Diaconis's coin tossing model, whether Pr(V and ω are in the black region at t_1| the coin is tossed at t_1) is objective. I have argued that data can support the claim that such probabilities have objective values in just the way that data can support the claim that mass values are objective. What is more, the value of this synchronic macro-probability can be viewed as the upshot of an earlier micro-variable C. Laplaceans thus find themselves in an untenable position – if micro-probabilities are objective, so too must macro-probabilities be objective.

Acknowledgements My thanks to Robert Brandon, Jeremy Butterfield, Juan Comesaña, Branden Fitelson, Patrick Forber, Malcolm Forster, Clark Glymour, Daniel Hausman, Thomas Hofweber, Marc Lange, Stephen Leeds, Hugh Mellor, Greg Novack, John Roberts, Alexander Rosenberg, Carolina Sartorio, Larry Shapiro, Christopher Stephens, Michael Strevens, and Jim Woodward for useful comments.

Notes

[1] With finite population size, fitness should not be defined as the mathematical expectation; see Sober (2001a) for discussion.

[2] Horan (1994) defends a similar position.

[3] Rosenberg (2001, p. 541) has more recently taken a different position: "… the way in which the principle of natural selection invokes the notion of probability cannot be understood either epistemically or in terms of probabilistic propensities of the sort quantum mechanics trades in. The notion of probability that the principle of natural selection invokes can only be understood as the kind of probability to which thermodynamics adverts [namely] … long run relative frequencies." My focus here will be on Rosenberg's earlier position, though I will have a little to say about both thermodynamics and the long run frequency interpretation of probability.

[4] In fact, there are reasons to think that Newtonian mechanics is not deterministic; see Earman (1986, 2004) and Butterfield (1998) for discussion.

[5] This is the view embraced by the theory of single-case propensities developed by Giere (1973).

[6] (MS) can be modified to accommodate the idea that some macro-properties supervene on historical facts, so that the supervenience base for macro-properties at time t is the state of particles in some temporal interval leading up to t.

[7] The following remark from Earman (2005) is a useful cautionary reminder to those who think that indeterminism is a settled matter in modern physics: "One might have hoped that this survey

[referring to his own paper] would provide an answer to the question: If we believe modern physics, is the world deterministic or not? But there is no simple and clean answer. The theories of modern physics paint many different and seemingly incommensurable pictures of the world; not only is there no unified theory of physics, there is not even agreement on the best route to getting one. And even within a particular theory – say, Quantum Mechanics or the General Theory of Relativity – there is no clear verdict. This is a reflection of the fact that determinism is bound up with some of the most important unresolved problems for these theories."

[8] Why does Figure 1 have an arrow going from B at t_2 to Y at t_2, while Figure 2 has a double arrow between (B_1 or B_2 or … or B_n) at t_2 and Y at t_2? The reason I drew Figure 2 in this way is that I wanted the demon to be able to derive a point value for Pr(Y at t_2|A at t_1) from the fact that Pr(B_1 or B_2 or … or B_n at t_2|A at t_1) = q. If the disjunction of the B_i at t_2 simply sufficed for Y at t_2, what would follow is just that Pr(Y at t_2|A at t_1) $\geq$ q.

[9] These two principles, L_1 and L_2, bear slightly different relationships to the question of which interpretation of probability you should use. Whether you are talking about subjective credences or objective chances, you still need advice about which conditional probabilities to use in making predictions. In this sense, L_1 is not wedded to any one interpretation of probability.

L_2, however, gives advice about which interpretation of probability you should impose on a given conditional probability.

[10] I am construing the principle of total evidence as saying, not just that you should use *all* of the evidence you have, but that using *more* of the evidence is better than using *less*.

[11] Why accept the principle of total evidence? Good (1967) constructs a decision-theoretic justification. He says that his argument applies to logical and subjective probabilities, but not to frequencies. I take the argument to apply to objective probabilities as well.

[12] The thesis that 1's and 0's are sticky asserts that
If $\Pr(A|B) = 1$, then $\Pr(A|B\&C) = 1$ (assuming that $\Pr(B), \Pr(B\&C) > 0$).

Proof. If $\Pr(A|B) = 1$, then $\Pr(A|B\&C)\Pr(C|B) + \Pr(A|B\&\text{-}C)\Pr(\text{-}C|B) = 1$ as well. If $\Pr(C|B) > 0$, then $\Pr(A|B\&C) = 1$. And $\Pr(C|B) > 0$, since $\Pr(B\&C) > 0$, $\Pr(B) > 0$.

[13] Strevens (2003) views (Macro SO Micro) as the key to understanding how there can be simple, stable, and objective probabilistic macro-laws; he calls this principle *the probabilistic supercondition*. Strevens is interested in cases in which it is exactly satisfied as well as in cases in which it is satisfied only approximately.

[14] Note that screening-off, as I define it, is not asymmetric; Micro's screening-off Macro from Y does not preclude Macro's screening-off Micro from Y. Screening-off is often defined as an asymmetric relation. Since my purpose in what follows is to argue that (Macro SO Micro) is rarely, if ever, true, my arguments, which are formulated in terms of the weaker notion, also count against the principle when the stronger interpretation is used.

[15] Although the equality asserted by (Macro SO Micro) can't be exactly correct when the macro-theory is probabilistic and the micro-theory is deterministic, it can be approximately true. Here's a simple case in which this is so. Suppose that the effect term Y at t_2 involves a statistic concerning outcomes over many trials. For example, what is the probability that a fair coin will land between 40% and 60% heads when tossed 1000 times? The probability is *nearly* one. The deterministic micro-details (supposing there are such) confer on that outcome a value that is equal to 1. So (Micro SO Macro) is approximately true.

[16] The monotonic increase of cancer risk with increase in the number of A particles inhaled is not essential for this argument. For each item in a set to have the same value as the set's average value, the items must all have the same value.

[17] Spirtes and Scheines (2003) discuss a similar example – Pr(heart attack|cholesterol level = α) will be an average, since there are different mixtures of high and low density lipids that instantiate the same cholesterol level, and different mixtures confer different probabilities on having a heart attack.

[18] The use of descriptors of the population's state at t_1 that are defined so that they entail a probability distribution for the system's state at t_2 is what leads mathematical models in population

biology to have the status of *a priori* mathematical truths (Sober 1984) and to be time-translationally invariant (Sober 1993c).

[19] It might be better to express this screening-off claim by absorbing the facts about viabilities at t_1 into the probability function, thus yielding the equality Pr_v(O is alive at t_2|O is alive at t_1) = Pr_v(O is alive at t_2|O is alive at t_1 & O has genotype G at t_1). That way, the conditional probabilities don't conditionalize on probability statements.

[20] The (Macro SO Micro) principle would not be rescued by allowing the times t_0, t_1, t_2 to be intervals rather than instants. As long as there is a temporal gap between t_1 and t_2, the kind of counterexample contemplated here can arise. Thus, even if the macro-state at t_1 supervenes on "historical" facts at the micro-level that extend back in time from t_1, the counterexample stands. And allowing the phenotype at t_1 to influence the phenotype that arises later won't save (Macro SO Micro), either.

[21] I say "may" because the question of whether G_1 at t_1 or P_1 at t_1 is a better predictor of whether Y is alive at t_2 depends on the quantitative relationships that obtain among the path coefficients we might associate with the arrows in the left-hand diagram in Figure 6.

[22] Although I have argued that (Macro SO Micro) is rarely if ever true when both the macro- and the micro-theories are probabilistic, there may be circumstances in which it is approximately correct. One of the main theses of Strevens (2003) is that (Macro SO Micro) is approximately true when the system is "macro-periodic" and approximately "micro-constant." The former says, roughly, that the probability distribution over initial conditions is smooth; the latter concerns how a given macro-state is realized by various micro-states. It suffices for micro-constancy if all the micro-state realizers of a given macro-state have the same probability of occurring, conditional on the occurrence of that macro-state.

[23] Woodward accepts the so-called causal Markov principle, which asserts that one of these three possibilities must be true; I do not (Sober 1988, 2001b). Woodward's ideas about intervention are interesting independent of this question.

[24] Kim (1989) argues that behavior cannot have both psychological and neurophysiological causes, since this would imply that behavior is over-determined, a consequence that Kim finds objectionable. However, this is not a case of overdetermination if overdetermination requires only that the two causes each be sufficient for the effect and *independent* (Kim 2000). When Holmes and Watson each shoot Moriarty through the heart at the same time, each cause could occur without the other. But if neurophysiological states (or those states plus relevant facts about the physical environment) provide supervenience bases for psychological states, the two will not be independent. A further objection to Kim's argument is relevant when the causal relation is probabilistic – there is no *over* determination if there is no determination at all.

[25] These brief remarks do not address the question of whether the micro supervenience base should be held fixed when the causal role of a macro-property is tested, or whether the macro-property should be held fixed when the causal role of its supervenience base is assessed; see Sober (1999b) for discussion.

[26] It may be objected that D and P_2 do not compete with each other as explanations of E_2 and that the *desideratum* of generality is relevant only when the task is to sort out competing explanations. If competing explanations must be incompatible, then I agree that "D occurs" and "P_2 occurs" are not competitors (and the same holds for "D causes E_2" and "P_2 causes E_2"); but then it follows that the micro- and macro-stories that Putnam describes in the peg-and-board example are compatible, so they are not competitors, either. On the other hand, if competition does not require incompatibility, I am unsure how this objection should be interpreted.

[27] This part of Nozick's argument derives from the indispensability arguments used by Quine (1953) and Putnam (1971) concerning the reality of mathematical entities. For a critique of these indispensability arguments, see Sober (1993a, 2000).

[28] I am not arguing that the existence of beliefs and desires (and of bowling balls, for that matter) is beyond dispute. Perhaps Patricia and Paul Churchland are right that beliefs and desires don't belong in a scientific psychology. My point is that this eliminativist conclusion is not supported by the fact that a hypothetical demon wouldn't need to postulate such things in order to make predictions.

[29] Here are two reasons for thinking that conditional probability should not be defined in terms of unconditional probabilities. First, there are cases in which it is perfectly clear that $Pr(Y|X)$ has a value even though $Pr(X) = 0$. Second, one sometimes knows perfectly well the value of $Pr(O|H)$ and that this is an objective quantity, even though it seems perfectly clear that $Pr(H)$ has no objective value at all; this situation often arises when H is a large-scale scientific theory (e.g., relativity theory), and O is some observational claim upon which the theory confers a probability. These points do not undercut the fact that $Pr(Y|X) = Pr(Y\&X)/Pr(X)$ when the two unconditional probabilities are well-defined.

[30] As further evidence against the claim that a proposition has one true probability, it is worth noting that, in an indeterministic process, the probability of Y at t will *evolve* (Sober 1984). For example, it is perfectly possible that $Pr(Y \text{ at } t_2|A \text{ at } t_1) \neq Pr(Y \text{ at } t_2|B \text{ at } t_0)$.

[31] This point about the actual relative frequency interpretation reveals a confusion in the Laplacean position. The fact that we shouldn't use $Pr(Y \text{ at } t_2|X \text{ at } t_1)$ to predict whether Y at t_2 if we can use $Pr(Y \text{ at } t_2|A \text{ at } t_1)$ instead doesn't show that the former has no objective interpretation. That conflates a question about pragmatics (which probabilities we should *use* to perform which tasks) with a question about semantics (Sober 2003a).

[32] In addition, $Pr(Y \text{ at } t_2|X \text{ at } t_1)$ will fail to represent the propensity of X at t_1 to cause Y at t_2 when Y at t_2 and X at t_1 are effects of a common cause at t_0. And even when X at t_1 has a causal impact on Y at t_2, the measure of this impact is not to be found in the value of $Pr(Y \text{ at } t_2|X \text{ at } t_1)$; the effect of X at t_1 on Y at t_2 is to be found by making some sort of comparison – e.g., between $Pr(Y \text{ at } t_2|X \text{ at } t_1)$ and $Pr(Y \text{ at } t_2|\text{notX at } t_1)$ when one controls for the other causally relevant properties of the situation at t_1. Causes are difference makers, so a single conditional probability does not represent a causal propensity.

[33] This definition of objectivity has the consequence that statements about someone's psychological state cannot be objectively true. To handle this type of case, it would be better to define objectivity by saying that believing the statement does not *make* it true (though puzzle cases would still remain, such as the proposition "I believe something"). These niceties won't matter for what follows.

[34] There is a difference between showing that our beliefs about mass are caused by something in the object and showing that the property of the object that does the causing is the object's *mass*. Because of space limitations, I'll glide over this distinction.

[35] Here I am adapting the common cause argument for realism that Salmon (1984) and Hacking (1985) defend. I do not claim that the anti-realist has no reply; rather, my point is that macro-probabilities are in the same boat as other theoretical quantities, like mass. This should be enough to show that there is no *special* objection to thinking that macro-probabilities are objective.

[36] AIC and other criteria of model selection impose penalties on a model for its complexity; see Burnham and Anderson (1998) for discussion. Model selection criteria permit one to answer the question of how much variation in the estimates made by different observers is consistent with regarding the common-cause model as better than the alternative model that postulates separate causes, one for each observer. The common-cause argument given here does not require a commitment concerning which model-selection criterion is best.

[37] A parallel claim is plausible concerning the objectivity of counterfactuals. Quine and others have noted that there are counterfactuals whose truth values are indeterminate (compare "If Verdi and Bizet were compatriots, they would be French" and "If Verdi and Bizet were compatriots, they would be Italian"). Some conclude from this that counterfactuals *never* describe objective matters of fact. I think this dismissal is too sweeping. For example, detailed counterfactuals that describe what would happen if you were to intervene in a causal system can be objectively true (Woodward 2003). For both conditional probabilities and counterfactuals, speakers often fail to be completely explicit, relying on shared information and context to supply relevant details.

[38] The Keller/Diaconis model does not assign probabilities to the initial conditions of a toss (its values for V and ω). It is not logically inevitable that the ratio of *areas* in the figure is the same as the ratio of *probabilities*.

[39] I treat the values of V and ω as "micro-descriptions" of a toss even though they apply to the whole coin toss apparatus without mentioning its parts. This is not objectionable since

the problem concerning the reality of macro-probabilities arises from the fact that they are "coarse-grained" – i.e., their conditioning propositions are less contentful than (complete) micro-descriptions would be.

[40] Here I am relying on the assumption that if n quantities are each objective, so is any quantity that is a function of their values; "mind-independence" evidently has this compositional characteristic.

[41] I argued in Section 5 that our desire for general explanations should not be taken to show that macro-probabilities are objective. In the present section, I have argued that we can have evidence that a macro-probability is objective when it provides a common cause explanation of various observations. This may seem inconsistent (eschewing explanation in one place, espousing it in another), but I think it is not. Philosophers have examined the concept of explanation from two angles. The problem that Hempel and his successors addressed is to decide which of various true propositions belong in an explanation of some target proposition. The problem addressed under the heading of inference to the best explanation is to say which of various candidate hypotheses one should regard as true. I don't regard the requirement of generality as having an objective status in Hempel's problem, but I do think there is an objective reason, in many cases, to view unified explanations as better supported than disunified explanations. See Sober (2003b) for discussion.

[42] Or rather, this is *an* answer, one constructed to appeal to Laplaceans. I don't rule out the possibility that $\Pr(\text{A at t1}|\text{X at } t_1)$ has the value it does because of a *macro* common cause, thus providing an instance of "downward causation."

[43] Notice that the decomposition of $\Pr(A|X)$ described here does not require that the common cause render the two effects conditionally independent of each other. In fact, it will not, if each A_i either entails X or entails not-X. Whereas the epistemic argument is naturally understood in terms of effects being unconditionally dependent though conditionally independent, the metaphysical argument should not be understood in this way.

References

Albert D (2000) Time and chance. Harvard University Press, Cambridge

Brandon R (1990) Organism and environment. Princeton University Press, Princeton, NJ

Brandon R, Carson S (1996) The indeterministic character of evolutionary theory – no 'no hidden variables proof' but no room for determinism either. Philos Sci 63:315–337

Burnham K, Anderson D (1998) Model selection and inference. Springer, New York

Butterfield J (1998) Determinism and indeterminism. In: Craig E (ed) The routledge encyclopedia of philosophy, vol. 3. Routledge, New York: pp. 33–39

Crane T, Mellor D (1990) There is no question of physicalism. Mind 99:185–206

Darwin C (1859) The origin of species. Harvard University Press, Cambridge, 1959

Dennett D (1987) True believers. In: The intentional stance. MIT Press, Cambridge

Diaconis P (1998) A place for philosophy? The rise of modeling in statistical science. Quart Appl Math 56:797–805

Earman J (1986) A Primer on Determinism. Kluwer, Dordrecht

Earman J (2005) Determinism – what we have learned and what we still don't know. In: Campbell J, O'Rourke M, Shier D (eds) Freedom and determinism, topics in contemporary philosophy series, vol II. Seven Springs Press. Also available at: www.ucl.ac.uk/∼uctytho/detearmanintro.html

Eells E (1981) Objective probability theory theory. Synthese 57:387–444

Forster M, Kryukov A (2003) The emergence of a macro-world: a study of intertheory relations in classical and quantum mechanics. Philos Sci 70:1039–1051

Garfinkel A (1981) Forms of explanation. Yale University Press, New Haven, CT

Giere R (1973) Objective single case probabilities and the foundations of statistics. In: Suppes P et al. (eds) Logic, methodology, and philosophy of science IV. North Holland, Amsterdam, pp 467–483
Glymour B (2001) Selection, indeterminism, and evolutionary theory. Philos Sci 68:518–535
Good IJ (1967) On the principle of total evidence. Br J Philos Sci 17:319–321
Gould S (1980) The Panda's thumb. W.W. Norton, New York
Hacking I (1985) Do we see through the microscope? In: Churchland P, Hooker C (eds) Images of science: Essays on realism and empiricism. University of Chicago Press
Hajek A (2003) What conditional probability could not be. Synthese 137:273–323
Horan B (1994) The statistical character of evolutionary theory. Philos Sci 61:76–95
Jackson F, Pettit P (1992) In defense of explanatory ecumenism. Econ Philos 8:1–22
Keller J (1986) The probability of heads. Am Math Monthly 93:191–197
Kim J (1989) The myth of nonreductive materialism. Proc Addresses Am Philos Assoc 63:31–47; reprinted in Supervenience and Mind. Cambridge University Press, Cambridge/England, 1993
Kim S (2000) Supervenience and causation – a probabilistic approach. Synthese 122:245–259
Laplace P (1814) A philosophical essay on probabilities. Dover, New York, 1951
Levi I (1967) Gambling with truth. Knopf, New York
Levi I, Morgenbesser S (1964) Belief and disposition. Am Philos Quart 1:221–232
Mayr E (1963) Animal species and evolution. Harvard University Press, Cambridge
Pearl J (2000) Causality – models, reasoning, and inference. Cambridge University Press, Cambridge
Putnam H (1971) Philosophy of logic. Harper, New York
Putnam H (1975) Philosophy and our mental life. Mind, language, and reality. Cambridge University Press, Cambridge/England
Quine W (1953) Two dogmas of empiricism. Harvard University Press, Cambridge, pp 20–47
Reichenbach H (1956) The direction of time. University of California Press, Berkeley, CA
Rosenberg A (1994) Instrumental biology or the disunity of science. University of Chicago Press, Chicago, IL
Rosenberg A (2001) Indeterminism, probability, and randomness in evolutionary theory. Philos Sci 63:536–544
Salmon W (1984): Scientific explanation and the causal structure of the world. Princeton University Press, Princeton, NJ
Skyrms B (1980) Causal necessity. Yale University Press, New Haven, CT
Sober E (1984) The Nature of Selection. MIT Press, Cambridge
Sober E (1988) The principle of the common cause. In: Fetzer J (ed) Probability and causation: essays in honor of Wesley Salmon. Reidel, Dordrecht, pp 211–28; reprinted in From a biological point of view. Cambridge University Press, 1994
Sober E (1992) Screening-off and units of selection. Philos Sci 59:142–152
Sober E (1993a) Mathematics and indispensability. Philos Rev 102:35–58
Sober E (1993b) Philosophy of biology. Westview Press, Boulder, CO
Sober E (1993c) Temporally oriented laws. Synthese 94:171–189; reprinted in From a biological point of view. Cambridge University Press, Cambridge, 1994, pp 233–252
Sober E (1999a) The multiple realizability argument against reductionism. Philos Sci 66:542–564
Sober E (1999b) Physicalism from a probabilistic point of view. Philos Stud 95:135–174
Sober E (2000) Quine's two dogmas. Proc Aristotelian Soc 74:237–280
Sober E (2001a) The two faces of fitness. In: Singh R, Paul D, Krimbas C, Beatty J (eds) Thinking about evolution: historical, philosophical, and political perspectives, vol 2. Cambridge University Press, Cambridge, pp 309–321
Sober E (2001b) Venetian sea levels, British bread prices, and the principle of the common cause. Br J Philos Sci 52:1–16
Sober E (2003a) Metaphysical and epistemological issues in modern Darwinian theory. In: Raddick G, Hodge MJS (eds) The cambridge companion to Darwin. Cambridge University Press, Cambridge, pp 267–287

Sober E (2003b) Two uses of unification. In: Stadler F (ed) The Vienna circle and logical empiricism – Vienna circle institute yearbook 2002. Kluwer, Dordrecht, pp 205–216
Spirtes P, Scheines R (2003) Causal inference of ambiguous manipulations. Philos Sci 71: 833–845
Spirtes P, Glymour C, Scheines R (2000) Causation, prediction, and search. MIT Press, New York
Stamos D (2001) Quantum indeterminism and evolutionary theory. Philos Sci 68:164–184
Strevens M (2003) Bigger than chaos. Harvard University Press, Cambridge
Tsai G (2004) Are higher level explanations always more general?" unpublished manuscript
Wilson R (2004) Boundaries of the mind. Cambridge University Press, Cambridge
Woodward J (2003) Making things happen. Oxford University Press, Oxford

Is Evolution An Optimizing Process?

James H. Fetzer

The thought that evolution invariably brings about progress deserves consideration, especially in light of the use of optimizing models of its operation. An extremely interesting collection of papers on evolution and optimality affords a suitable point of departure. In *The Latest on the Best* (1987), for example, some who believe that evolution should be viewed as an optimizing process join issues with others who do not. As the editor observes,

> *A number of the contributors from the earlier parts of the anthology ... argue forcefully for the difficulties with any general assumption that evolution generates optimal adaptation. On the other hand, a number of the contributors involved in applications of optimality analysis are extremely enthusiastic about the potential benefits of this approach (Dupre 1987, p. 22).*

Indeed, as he also observes, even within biology itself there appears to be more enthusiasm for applications of optimality theory than there is for the conclusion that evolution itself is an optimizing process (Dupre 1987, p. 22).

It may be useful here to approach this issue by exploring the differences between optimizing and satisficing approaches toward understanding evolution. In the process, it will be important to deal with the claim that satisficing has no clear meaning in evolutionary theory. By focusing upon the probabilistic character of fitness, however, it ought to be possible to shed light upon these difficult problems, since neither the frequency interpretation nor the propensity interpretation of fitness supports the conception of evolution as an optimizing process. The general conception that emerges from these reflections is that selection takes place relative to adaptations that are "good enough" to sustain survival and reproduction. While optimizing theory can play a valuable heuristic function in suggesting hypotheses about the design of organisms, satisficing theory clearly affords a more adequate framework for understanding the broadest and deepest features of evolution as a process.

J.H. Fetzer (✉)
McKnight Professor Emeritus, University of Minnesota Duluth, 800 Violet Lane, WI 53575, Oregon
e-mail: jfetzer@d.umn.edu

E. Eells and J.H. Fetzer (eds.), *The Place of Probability in Science*, Boston Studies in the Philosophy of Science 284, DOI 10.1007/978-90-481-3615-5_7,

The Tautology Problem

The issues encountered within this domain are closely related, if not identical, to those that revolve about the suitability of the phrase, "the survival of the fittest", as a general depiction of the evolutionary process. As Ernst Mayr observes, Charles Darwin borrowed the phrase from Herbert Spencer (Mayr 1982, secs. 386 and 519, for example). After all, if those who survive are the fittest, then the process of natural selection would appear to produce increasingly fit organisms across time, where there can be no basis for doubting whether the latest must be the best. If biological evolution is a process that invariably induces the survival of the best, then it must be impossible for the lastest not to be the best.

Although May is concerned that the phrase, "the survival of the fittest", may be merely tautological as a claim that cannot be false by virtue of its meaning, when properly understood, it not only appears to be no tautology but may even be empirically false. The problem is to separate the meaning of "fitness" from the meaning of "those who survive", which may be accomplished by viewing "fitness" as a probabilistic tendency to survive and reproduce. (A closely related view is proposed by Sober (1984, Ch. 2).) When the exact character of this probabilistic tendency is rendered more precise, however, it then becomes evident that "high fitness" may or may not be associated with increased frequency for survival and reproduction.

Optimizing Versus Satisficing I

Although we shall pursue more exact characterizations in the following, it may be worthwhile to provide a general sketch of the differences which tend to distinguish "optimizing" from "satisficing". These notions are derived from decision theory, where optimizing models characterize systems that always select solutions to problems that are at least as "good" as any other solution; satisficing models, by comparison, characterize systems that select solutions to problems that are "good enough" but which may yet have alternatives that are even better. Judgments of goodness (good, better, and best), of course, are relative to some presupposed set of values or utilities. (Michalos (1973) provides an illuminating introduction to the alternatives.)

When these ideas are applied within the context of evolutionary theory, however, an important distinction must be drawn. Decision theory is normative and prescribes how people should act, whereas evolutionary theory is explanatory and describes how nature does work. While normative theories do not have to be abandoned or repaired when they do not describe the way things are, explanatory theories in science have to be abandoned or repaired when they do not describe the way things are. If nature does not operate in conformity with optimizing models, therefore, then that counts against their standing as scientific theories, although optimizing models for decision making are not thereby disqualified, even if no one acts in conformity with them.

Evolutionary Values or Utilities

The values or utilities that are commonly assumed to possess evolutionary significance, of course, are those of survival and reproduction, especially offspring production equal to or greater than replacement level. Relative to this measure of utility, organism x may be supposed to have "higher fitness" than organism y, with respect to a specific environment, for example, when x has greater probability of offspring production than y, with respect to that environment. As Elliott Sober has observed, there are good reasons to resist the temptation of identifying "fitness" in this sense with actual reproductive success, since the property appears to be best envisioned as a disposition toward reproduction in lieu of its actual obtainment (Sober 1984, pp. 43–44).

Since nature cannot be expected to select solutions to problems that are unavailable, whether or not evolution should be viewed as an optimizing or as a satisficing process must be relative to the available gene pool and the available environment rather than relative to every possible gene pool and every possible environment. If these stronger conceptions are taken to describe kinds of "global optimality", then only weaker conceptions of "local optimality" deserve consideration here. Evolution would appear to qualify as an optimizing process in the appropriate sense, therefore, so long as selection produces organisms with fitness values that are equal to or greater than those of parent generations across time, but otherwise as satisficing.

Selection Is Not "Single-Step"

To ascertain whether or not evolution should be understood as an optimizing or as a satisficing process, therefore, it is essential to secure the right kind of non-tautological connection between "fitness" and "those who survive" and then to determine if the resulting process is optimizing, satisficing, or something else instead. If the less fit sometimes survive and reproduce while the more fit do not, for example, then surely that is something that evolutionary theory ought to be able to explain. Ultimately, I believe, evolution needs to be understood by means of the repetitive operation of "single case" tendencies for particular traits to be selected under specific (possibly unique) conditions utilizing the idea of single-case propensities.

This conception, however, should not be confused with the notion of "single-step" selection in the sense that Richard Dawkins has introduced (Dawkins 1986; reprinted in Ruse 1989, pp. 67–68). Single-case tendencies are compatible with cumulative selection operating across time, but single-step selection is not. Indeed, as Dawkins himself has observed,

> *There is a big difference ... between cumulative selection (in which each improvement, however slight, is used as a basis for future building), and single-step selection (in which each new "try" is a fresh one).*

If evolution had to rely on single-step selection, it would never have got anywhere (Dawkins 1986 as reprinted in Ruse 1989, p. 68).

The application of the single-case propensity interpretation to the problem at hand, therefore, should not be rejected on the ground that evolution is a cumulative rather than a single-step process. That would be unjustifiable.

The Nature of Optimality

In their valuable paper on optimality theory in evolutionary biology, G. A. Parker and J. Maynard Smith, both noted for their contributions to this domain, consider the question, "Can natural selection optimize?" In their view, optimization by natural selection requires special conditions:

> *The ability of natural selection to optimize depends on gene expression and on the mechanism of genetic change in populations. Of particular importance are the rate at which selection can alter the genetic structure, the amount of additive genetic variance present at the start of selection, gene flow (for example, that arising from immigration), the rate at which conditions change, and random effects such as genetic drift. Most optimality models assume that strategies reproduce asexually, or if the model is Mendelian, that the optimal prototype can breed true. Pleiotropy (genes affecting multiple traits) is assume not to operate and strategies are allowed to replicate independently of each other. Obviously selection cannot produce an optimum if there is no way of achieving it genetically, but for some models, it is clear that selection will get as close to the optimum as the genetic mechanism will allow (Parker and Maynard Smith 1990, pp. 30–31).*

When natural selection occurs in asexually reproducing or true breeding populations within stable environments that endure unchanged for periods of time which are sufficient for optimal traits to emerge, optimizing models of evolution may be appropriate in the absence of pheiotropic effects. In situations of the kind that characterize most natural as opposed to artificial environments, however, conditions like these are not realized.

Indeed, even Parker and Maynard Smith themselves observe that the emergence of optimal traits is not easily arranged: "infinite time and infinite populations would be needed to achieve the [evolutionary] peak itself" (Parker and Maynard Smith 1990, p. 31). But this means that optimality theory provides an idealized conception of what might happen in the limit as a special case rather than a descriptive explanatory framework for understanding evolution under normal conditions. A satisficing model, by comparison, provides a foundation for viewing the emergence of optimal adaptations as a possible "long run" product of what has to be understood as a "short run" process that applies to finite populations and finite times.

An Alternative Framework

The tendency in evolutionary theory to fixate on the long run rather than on the short run or on the single case has a parallel in the theory of probability itself, where long-run frequency conceptions have prevailed until recently. The emergence of the single-case propensity conception within this context thus promises to shed light on natural selection. Indeed, from the perspective afforded by the single-case propensity conception, appeals to the ultimate outcome of long-run processes appears to have the general character of teleological causation, whereas appeals to the proximate outcomes of single-case processes appears to have the general character of mechanistic causation instead (e.g., Fetzer 1988).

If optimal adaptations only emerge as a special limiting case under quite idealized conditions, moreover, it should be evident that the products of evolution that emerge during any merely finite segment of the world's history are never optimal, unless allowance is made for their production as a consequence of fortuitous conditions (or by chance). But if this is indeed the case, then satisficing appears to be the strongest theory that generally applies. While Parker and Maynard Smith consider "the comparative method" and "quantitative genetics", they do not consider satisficing approaches at all. Once the properties of satisficing models are properly understood, however, their appeal should become obvious.

The Probabilistic Connection

Evolution appears to be an inherently probabilistic process, at least to the extent to which sexual selection, sexual reproduction, genetic drift and the like involve probabilistic elements. Two interpretations of probability might apply here, the long-run frequency interpretation, which identifies probabilities with limiting frequencies in infinite sequences, and the single-case propensity conception, which identifies them with the strengths of causal tendencies to bring about particular outcomes on individual trials. While we have already discovered that optimal adaptations as a result of natural selection only occur (other than by chance) over the long run, perhaps evolution can still be viewed as a probabilistically-optimizing process.

Suppose we define "fitness" in terms of probability of survival and re-production at a level equal to or greater than replacement level R within the specific environment E. Thus, when $P(Ri/Ei \,\&\, i = x) > P(Ri/Ei \,\&\, i = y)$, x is more fit than y with respect to R in E. Then it might be plausible to hold that less fit traits will decrease in frequency across time or that less fit traits will eventually no longer be "good enough" to sustain survival and reproduction. But when such claims are subjected to close scrutiny in relation to the frequency and the propensity alternatives, it becomes clear that they cannot be sustained. Even when claims about what will probably occur displace claims about what will occur, these probabilistic alternatives do not support the idea of evolution as an optimizing process.

The Frequency Interpretation

On the frequency interpretation of probability, for example, the probability for B in relation to A equals p – that is, $P(B/A) = p$ – if and only if the limiting frequency for B in an infinite sequence of A trials equals p. The relative frequency for B in finite segments of A trials can arbitrarily vary from p. Whenever $P(B/A) = p$, of course,

$P(-B/A) = 1 - p$. Even if p is high and $1 - p$ is low, therefore, outcomes with low probability can occur, even with high relative frequency, across any finite segment of the world's history. While the frequency interpretation guarantees convergence between probabilities and frequencies over the long run, it does not guarantee convergence of probabilities and frequencies in the short run.

If we identify outcomes of kind B with offspring production at equal to or greater than replacement level relative to environments of kind A, then it should be apparent that, even when that outcome has high probability, the result of offspring production at less than replacement level may still occur, over the short run, in relation to the frequency approach. Moreover, if we identify outcomes of kind B instead with the production of offspring whose own fitness values are equal to or greater than those of their parents, then it should still be apparent that, even when that outcome has high probability, the result of the production of offspring whose own fitness values are less than those of their parents may likewise occur.

The Propensity Interpretation

On the propensity interpretation of probability, matters are even worse, since, as a property of each single trial of those conditions, even infinitely many repetitions of those conditions cannot guarantee that an outcome will occur with a limiting frequency that equals its generating propensity. This interpretation maintains that the probability for B in relation to A equals p if and only if there is a causal tendency of strength p for conditions A to produce (or "bring about") an outcome of kind B on any single trial of that kind. The conditions that specify a trial of kind A can be broadly construed to include conceptions and gestations that endure over intervals of time so long as every property that matters to survival and reproduction is considered.

The ontological difference between the frequency and the propensity interpretations is that one makes probabilities properties of infinite sequences while the other makes them properties of singular trials. The advantage of the propensity approach, from this point of view, is that, if the world's history is merely finite, the existence of probabilities as propensities remains secure, while the existence of probabilities as frequencies is problematical. The logical connection that obtains between probabilities and frequencies on the frequency approach, however, is completely severed by the propensity approach, because, even over an infinite sequence of trials, the limiting frequency for outcome B might deviate arbitrarily from the propensity for B.

Reviewing the Argument

Three possible foundations for the conception of evolution as an optimizing process have been considered, none of which provides suitable support. The first came from Parker and Maynard Smith's discussion of the conditions that are required for natural selection to produce optimal adaptations. When natural selection occurs in asexually reproducing or true breeding populations within stable environments that endure unchanged for periods of time that are sufficient for optimal traits to emerge, optimizing models of evolution can be applied in the absence of pheiotropic effects. Most natural situations do not satisfy the conditions, however, and optimal adaptations require infinite populations and time.

This means that optimality theory provides an idealized conception of what might happen in the limit as a special case under fixed conditions rather than a descriptive explanatory framework for understanding evolution under ordinary conditions. Moreover, appealing to probabilistic properties does not appear to salvage the situation. Under the frequency interpretation, higher fitness yields higher frequencies of survival and reproduction (or higher fitness in offspring generations) only by assuming infinite sequences of trials. Under the propensity interpretation, higher fitness affords no guarantee of higher frequencies of survival and reproduction (or higher fitness in offspring generations) even if infinite sequences are assumed.

Random and Accidental Factors

Probabilistic factors are an extremely important source of difficulty for the conception of evolution as an optimizing process, especially in relation to the mechanism of natural selection, but they are not the only source of difficulties. Others include different ways in which various environments can be subject to change, which shall be referred to here (somewhat arbitrarily) as "random happenings" and as "accidental occurrences" as follows:

(i) Random happenings (such as stray bullets, terrorist bombs, and the AIDS virus) are micro properties which vary within macro environments. None of us has greater fitness in coping with them as a rule, but only some of us actually encounter them. Organisms with high fitness within macro environments may not survive and reproduce due to their influence, while organisms with low fitness within those macro environments may instead.

(ii) Accidental occurrences (including large asteroids hitting the Earth) are events that bring about major alterations in the macro environment, which we otherwise tend to treat as a "closed system". Once again, organisms with high fitness within macro environments may not survive to reproduce due to their influence, as may have happened with the dinosaurs. Other organisms of lower fitness, however, might nonetheless survive, etc.

Optimizing Versus Satisficing II

The point of introducing these concepts is not to suggest that evolutionary theoreticians have been oblivious to the role of random factors or of accidental occurrences as they have been defined above (which Arnold and Fristrup 1982, among others, have acknowledged), but rather that these are other elements that further complicate the conditions that must obtain for optimizing models to obtain. If the conditions under which natural selection operates are constantly changing due to the influence of random factors and of accidental occurrences, then reliance upon models that presuppose infinite populations across infinite times relative to unchanging environments cannot possibly provide an adequate foundation for describing or explaining the actual course of the evolutionary process.

Perhaps the most important difference between optimizing and satisficing theory, from this perspective, is that optimizing theory implies that, over the long run, if not the short, organisms become increasingly more and more fit as a result of an inevitable "winnowing process" that takes place within constant environments. Satisficing theory, however, carries no such implication, accepting instead that sometimes higher fitness may not increase in relative frequency across time, while lower fitness does. The satisficing approach, nevertheless, provides a foundation for viewing the emergence of optimal adaptations as a possible "long run" product of what has to be understood as a "short run" process that applies to finite populations and finite times, under certain highly fortuitous conditions.

Fitness as a Propensity

Other authors have proposed that the propensity interpretation of probability might provide an appropriate foundation for understanding the nature of fitness, especially Susan Mills and John Beatty (1979). Their conception, however, does not properly represent either the context-dependence of propensities or their character as single-case causal tendencies. Consequently, they encounter difficulty in developing an adequate account of fitness as a propensity, which leads them to reconsider the adequacy of this approach (Beatty and Feinsen 1989). The difficulties they consider, however, appear to be problems with the notion of fitness rather than problems with the notion of propensities, when that conception is properly understood (cf. Fetzer 1986).

Some of the most important reservations that have been raised concerning propensities, moreover, concern their formal properties. Beatty and Feinsen (1989), for example, assume that propensities are properly formalized as conditional probabilities and that there must be a positive correlation between high fitness and reproductive success. These assumptions, which others, including Alexander Rosenberg and Robert Brandon, have endorsed, provide a misleading characterization of the propensity interpretation and generate problems that are more apparent than real (cf. Fetzer 1981 and Niiniluoto 1988). Other objections are also defeated by the failure to appreciate that any statistical measures constructed on the basis of the propensity interpretation qualify as propensity concepts.

Permanent and Transient Properties

Since fitness is relative to environments and, on probabilistic interpretations, characterizes the strength of tendencies to survive and reproduce in those environments by organisms that differ in their properties, it can be tempting to attempt to reduce fitness to those properties. In *Adaptation and Environment* (1990), for example, Robert Brandon identifies dispositions with their underlying physical bases. To be water soluble "just means" to have a certain molecular structure; to have a specific probability of coming up heads on a toss "just means" having a certain physical structure; and so forth. This promotes the identification of fitness as a disposition with specific phenotypes, where the specific behavior organisms display that matters to selection tends to be ignored.

This approach, however, overlooks that the same dispositions might be possessed by different kinds of things and that dispositions tend to have multiple manifestations. Water has the same molecular structure, even when it is found in solid, liquid, and gaseous states, where the behavioral abilities of water molecules in those different states differ significantly. Coins of different sizes and shapes can still be "fair", even though they have different physical structures. A distinction has to be drawn between reference properties of things like water molecules and coins and the dispositions they have that are permanent – which they cannot be without, even though they are non-definitional – and transient – which they can be without, while remaining things of the same kind. It is wrong to reduce dispositions to reference properties (Fetzer 1981, 1986).

Laws and Levels

The distinction has multiple ramifications. A slot machine that could be loaded with a deterministic random-number generator or an indeterministic causal mechanism could yield the same results "by chance", yet the outcomes of the first would be predictable without exception, where those of the second were only probabilistically expectable. Since these are components that are exchangeable, the kind of outcome behavior the system displays is a transient rather than a permanent property. Whenever the description of the machine includes the number random-number generator or the indeterministic mechanism, however, then that mode of behavior would be permanent rather than transient, not by definition but because those kinds of response outcomes are themselves permanent properties of mechanisms of those different kinds, respectively.

Similarly, it wrong to reduce the response outcomes of species to their phenotypes, even though, for the lower species, every member of that species may have the same behavioral dispositions among its permanent properties. Once we recognize that, for the higher species, different behavior responses can occur as a consequence of learning (in the form of classical conditioning, operant conditioning, and imitative learning, for example, not to mention the role of thinking and of reasoning in affecting behavior), at least some of which are indeterministic, even under the same external conditions, it should be clear that, even if genes are the units of selection, the level of selection is not that of the phenotype but of behavior, which succeeds or fails (Fetzer 2005).

A Fourth Argument

While I would like to think I have already provided sufficient evidence that evolution is not an optimizing process, I want to advance at least one additional argument that is meant to make it very obvious that there are evolutionary phenomena that cannot be appropriately understood on the basis of an optimizing model but which make good sense from the point of view of a satisficing model. This case has been offered by David S. Wilson (1980), who suggests that there is nothing about evolutionary theory that precludes the possibility that selection might "routinely" occur under conditions whereby an organism x decreases its own fitness but nevertheless is selected because it has decreased the fitness of competing conspecifics y even more (Sloan 1980 as reprinted in Brandon and Burian 1984, p. 275).

An appropriate illustration would appear to be American political campaigns, where negative advertising is normally judged successful as long as it wins more votes than it loses. There is an instructive lesson for evolutionary theory here, moreover, since cases of this kind emphasize that selection tends to be a matter of relative rather than of absolute fitness, insofar as those traits that confer a competitive advantage with respect to survival and reproduction can be "good enough" even if better solutions are available. If organisms can benefit from diminishing their fitness in order to increase their relative advantage over others, then it is difficult to see how the idea of evolution as an optimizing process can be justified.

The Role of Optimizing Theory

It could be maintained that, even if natural selection (or evolution, which is not confined to the mechanisms of natural selection) is not an optimizing process, optimality theory remains a useful heuristic device for generating hypotheses about phenotypic design, even if "not about the process of evolution that produced that design" (Smith 1990b). In suggesting this alternative, Eric Smith admits that there may be significant philosophical questions that can be raised about using optimality theory to study something, such as phenotypic design, that is the result of a non-optimizing process, but he cites the publication of thousands of research articles in which optimality theory fulfills precisely such a role.

In advancing this position, Smith would seem to echo the findings of John Dupre, when he reports that within biology itself there appears to be more enthusiasm for applications of optimality theory than exists for the conclusion that evolution itself is an optimizing process (Dupre 1987, p. 22). And, indeed, the potential for optimality theory to fulfill a role in generating hypotheses about degrees of adaptiveness of phenotypes relative to specific environments does not contradict or undermine the view that fitness should be properly understood as a single-case dispositional property, where a clear distinction is drawn between fitness values and reproductive success. Optimality theory can be useful in a heuristic role.

Instrumentalism Versus Realism

The question of whether evolution is an optimizing process, therefore, has to be distinguished from the question of whether optimality analysis can be useful in generating fitness hypotheses. Treating natures "as if" it were an optimizing process may be beneficial as a heuristic technique, but that does not mean that evolution itself is an optimizing process. The issue here is closely related to the distinction between realism and instrumentalism as it arises in theorizing generally. Instrumentalism views theories as instruments of prediction that are not meant to describe entities or properties in the world, while realism views theories as descriptions of the world. It seems evident that optimality theory can be useful instrumentalistically.

Thousands of published handbooks of navigation being with a sentence to the effect, "For present purposes, we will assume that the Earth is a small stationary sphere whose center coincides with that of a much larger rotating stellar sphere", as Thomas Kuhn has observed (Kuhn 1957). The wide-spread utility of adopting a certain model for a special purpose, however, in no way alters the limitations of that model for understanding the world itself. If evolution is not an optimizing process, then it makes no difference that, for certain special purposes, we can treat it "as if" it were. As long as our purpose is to understand the nature of evolution, therefore, optimizing theory is not enough. Its value appears to be exclusively heuristic in kind.

Is Natural Selection Optimizing?

Smith has also endorsed the fall-back position that, even if evolution is not an optimizing process, natural selection may still remain as an optimizing component of a complex process. When consideration is given to the many kinds of factors that contribute to the evolutionary process, including genetic mutation, sexual reproduction, genetic drift, natural selection, sexual selection, group selection, and artificial selection, it appears difficult to believe that the complex interaction of these factors ought to produce increased fitness in offspring reproduction across time. And the idea that natural selection as such might still qualify as an optimizing component of this complex evolutionary process does not seem to fare better.

The phrase, "natural selection", of course, can be taken as synonymous with the evolutionary process as a process of selection by nature, or it can be viewed as one component of a more complex process. Either way, however, every argument that we have considered here still applies. If actual environments are subject to change (due to the influence of random happenings and of accidental occurrences, for example) and do not involve infinite populations and infinite time, then natural selection cannot produce optimal adaptations except under fortuitous conditions, as explained above. Probabilistic variations do not make matters any better, moreover, where the most promising account makes optimal adaptations a possible outcome of the short-run operation of what must be viewed as a single-case process.

Is Evolution an Optimizing Process?

The strongest arguments supporting the conception of evolution as an optimizing process thus appear to be indefensible. The classic model advanced by Parker and Maynard Smith depends on special conditions that are seldom, if ever, satisfied during the history of world, including, for example, infinite populations and infinite times. Two probabilistic models are available on the basis of the frequency and the propensity interpretations. The frequency-based model, however, can only guarantee that higher fitness will produce higher reproductive success over the infinite long run. The propensity-based model does not even guarantee the convergence of fitness and success over the infinite long run.

These considerations suggest that evolution should not be viewed as an optimizing process. Sometimes higher fitness occurs with higher frequency across time, but sometimes not. The role of random factors and accidental happenings reinforces the variability of environments, which by itself undermines the applicability of optimality models. Nevertheless, optimality theory does appear to be applicable in the heuristic role of suggesting hypotheses about the adaptiveness of phenotypic design in generating fitness hypotheses. Taken altogether, therefore, the question of whether evolution should be viewed as an optimizing process seems to have a definite answer, which deepens our understanding of nature.

Is Evolution a Satisficing Process?

Whether or not evolution should be viewed as an optimizing process, as a satisficing process or as something else instead, as suggested above, depends upon the mechanisms that bring evolution about. If optimizing models characterize systems or processes that always select solutions to problems that are at least as "good" as any other solution, while satisficing models characterize systems or processes that select solutions to problems that are "good enough" but can still have alternatives that are even better, then even if evolution is clearly not an optimizing process, it might not be a satisficing process, either. Perhaps evolution is something else entirely.

Indeed, there is something to this objection. If those who survive are not necessarily the fittest, the process of natural selection might or might not produce increasingly fit organisms across time. Since biological evolution is not a process that invariably induces the survival of the best, it remains entirely possible that the latest are not the best. But insofar as natural selection and biological evolution do occur even in the absence of optimal strategies and optimal adaptations, it appears as though nature is content with genetic combinations and selection processes that are simply "good enough". Selection and evolution operate relative to available genes and available environments that satisfy their own non-optimal conditions.

Evolution as a Gamble with Life

To avoid any misunderstanding of the nature of my argument, bear in mind that evolution has several distinct dimensions. It can be viewed as a causal process involving interaction between organisms with different degrees of fitness and their environments. Since more fit organisms, by hypothesis, have higher probabilities of survival and reproduction, given a probabilistic conception of fitness, it would be mistaken to suppose that I am denying something that cannot be false. More fit organisms definitely do have high probabilities of survival and reproduction, even if, as we have found, there is no guarantee that the traits of organisms with higher fitness will invariably increase in their relative frequencies across time.

The actual course of evolution that emerges across time depends upon an interaction between organisms and their environment, where the environment is almost constantly changing and selection is a function of the behavior that they display (Fetzer 2005, 2007). Indeed, there are at least three reasons why environmental variations tend to defeat the emergence of optimal adaptations. Because actual environments only remain constant over finite intervals, not for infinite durations; because random happenings and accidental occurrences are exerting their influence on the course of evolution; and because those who survive and reproduce may be merely the lucky rather than the fit, the process of evolution across time appears to be but a gamble with life where players shift from game to game without any advanced warning.

Mechanism Versus Teleology

Ultimately, I believe, appeals to optimizing theory within evolutionary biology ought to be recognized as lingering remnants of a teleological metaphysics that a more thoroughgoing mechanistic conception can do without. There is no need to remain fixated on infinite populations and times or on interpretations of probability which guarantee some long-run convergence between probabilities and frequencies. What is required is a conception of probability as a single-case propensity that can be applied no matter whether the world's history is short or is long and without regard for either the size of the available population or the duration of the process of evolution.

The reasons for preferring the satisficing conception to the optimizing conception seem to be essentially the same as those for preferring a mechanistic to a teleological worldview. The history of science tends to reflect shifts from a teleological worldview (associated with Aristotle) to a mechanistic worldview (associated with Newton), on the one hand, and from a deterministic worldview (associated with classical mechanics) to an indeterministic worldview (associate with quantum mechanics), on the other. Abandoning the last vestiges of teleology in biology may appear to be a threat to a familiar worldview. But it can be replaced by another world-view that integrates evolution within an even more scientific framework.

Acknowledgements This paper had its origin in an email exchange with Eric Smith and Kim Hill, to whom I am indebted for commenting on my position. The original communications between us are cited for reference as Fetzer (1991a, b, c), Hill (1991a, b), and Smith (1991a, b). This is a somewhat revised version of Chapter 2 of Fetzer (2007).

References

Arnold AJ, Fristrup K (1982) The theory of evolution by natural selection: a hierarchical expansion. Paleobiology 8:113–129

Beatty J, Finsen S (1989) Rethinking the propensity interpretation: a peek inside Pandora's box. In: Ruse M (ed) What the philosophy of biology is – essays for David Hull. Kluwer, Dordrecht, The Netherlands, pp 17–31

Brandon RN (1990) Adaptation and environment. Princeton University Press, Princeton, NJ

Brandon RN, Burian RM (eds) (1984) Genes, organisms, populations: controversies over the units of selection. MIT Press, Cambridge, MA

Dawkins R (1986) The blind watchmaker: why the evidence of evolution reveals a universe without design. W. W. Norton, NY

Dupre J (ed) (1987) The latest on the best. MIT Press, Cambridge, MA

Fetzer JH (1981) Scientific knowledge. D. Reidel, Dordrecht/Holland

Fetzer JH (1986) Methodological individualism: singular causal systems and their population manifestations. Synthese 68:99–128

Fetzer JH (1988) Probabilistic metaphysics. In: Fetzer J (ed) Probability and causality. D. Reidel, Dordrecht/Holland pp 109–132

Fetzer JH (1991a) Thesis: evolution is not an optimizing process. HBES EMAIL NEWSLETTER, Internet 6 November 1991, 09:21:00 EDT

Fetzer JH (1991b) Survival of the fittest: a response to Eric Smith and Kim Hill. HBES EMAIL NEWSLETTER, Internet 12 November 1991, 08:56:00 EDT

Fetzer JH (1991c) Probabilities and frequencies in evolutionary theory: response to the replies of Eric Smith and Kim Hill. HBES EMAIL NEWSLETTER, Internet 19 November 1991, 13:47:00 EDT

Fetzer JH (2005) The evolution of intelligence: are humans the only animals with minds? Open Court, Chicago, IL

Fetzer JH (2007) Render unto Darwin: philosophical aspects of the Christian right's crusade against science. Open Court, Chicago, IL

Hill K (1991a) Re Fetzer on optimizing vs. satisficing. HBES EMAIL NEWSLETTER, Internet 7 November 1991, 09:58:36 EST

Hill K (1991b) From Kim Hill. HBES EMAIL NEWSLETTER, Internet 13 November 1991, 09:18:00 EST

Kuhn TS (1957) The Copernican revolution. Harvard University Press, Cambridge, MA

Mayr E (1982) The growth of biological thought. Harvard University Press, Cambridge, MA

Michalos A (1973) Rationality between the maximizers and the satisficers. Policy Sci 4:229–244

Mills S, Beatty J (1979) The propensity interpretation of fitness. Philos Sci 46:263–286

Niiniluoto I (1988) Probability, possibility, and plenitude. In Fetzer J (ed) Probability and causality. D. Reidel, Dordrecht/Holland, pp 91–108

Parker GA, Maynard Smith J (1990) Optimality theory in evolutionary biology, Nature 348:27–33

Ruse M (ed) (1989) The philosophy of biology. Macmillan, NY

Smith E (1991a) Re Fetzer on optimizing vs. satisficing. HBES EMAIL NEWSLETTER, Internet 6 November 1991, 11:38:17-0800 PST

Smith E (1991b) Rebuttal to Fetzer's reply. HBES EMAIL NEWSLETTER, Internet 13 November 1991, 16:25:18-0800 PST

Sober E (1984) The nature of selection: evolutionary theory in philosophical focus. MIT Press, Cambridge, MA

Wilson DS (1980) The natural selection of populations and communities. Benjamin/Cummings, Menlo Park, CA

Part III
Probabilities as Explanatory Properties

Propensity Trajectories, Preemption, and the Identity of Events

Ellery Eells

Introduction

The problem of preemption for theories of causation is well known. In its original and basic form, it is a problem for theories of causation that take a cause to be some sort of a necessary condition for its effects (e.g., an INUS condition as in Mackie (1974), or an event upon which an effect counterfactually depends as in Lewis and David (1973, 1986). The problem is that an event c may very well be a cause of an event e even though, had c not occurred, a "backup" event c' which also occurred would have caused e instead; in this case, c is not necessary for e, and c "preempted" c' in causing e. A clear and simple example of this is described by Lewis (2000) and Hall (2004). Suzy and Billy both want to break a bottle with a rock; they are both experts at this; they both pick up rocks and throw, accurately; Suzy's rock reaches the bottle first, breaking the bottle; by the time Billy's rock arrives the bottle is already shattered and his rock hits no glass at all. In this case it is clear that it was Suzy's throw (c) that caused the bottle to break (e), but Billy's throw (c') would have caused the bottle to break had Suzy not thrown her rock. In this case, Suzy's throw was not, under the circumstances, necessary for the bottle to break; her throw preempted Billy's throw in the breaking of the bottle.[1]

Recently, the problem of preemption has been applied to *probabilistic* theories of causation, on which a cause need not be necessary for its effects, but only raise the probability of its effects. Menzies (1996) has applied the problem to counterfactual theories of probabilistic causation, and Ehring (1994, 1997) has applied the problem to a "probability (or propensity) trajectory" approach to probabilistic causation that I have earlier urged (1991). In this paper, I will be concerned with applying this trajectory approach to cope with the problem of (what I will call) probabilistic preemption. I begin in Section 1 with a summary of the trajectory theory. In Section 2, I elaborate, in the context of the trajectory theory, an approach to the problem of preemption that I favor and that has been previously suggested in the context of counterfactual theories of causation but whose versatility I think has not been fully appreciated. This will involve a certain conception

E. Eells (✉)
University of Wisconsin, Madison, WI, USA

E. Eells and J.H. Fetzer (eds.), *The Place of Probability in Science*, Boston Studies in the Philosophy of Science 284, DOI 10.1007/978-90-481-3615-5_8,

of events and a substantive thesis concerning events so conceived. In Section 3, I discuss the possibility that thesis might be false, and also discuss the more general phenomenon of *overdetermination* (or *redundancy*, of which preemption, or *asymmetric* overdetermination, is a species).

Trajectories

In this section I briefly summarize the probability (or propensity) trajectory theory of singular probabilistic causation.[2] This will be done by noting just five features of the theory. First, it is a theory of singular causation. Thus, it seeks to understand such "token level" claims as "Harry's smoking caused his heart attack" or "Suzy's throwing the rock (at that time and place) caused that bottle to break", rather than such "type level" claims as "Smoking is a positive causal factor for heart attacks" or "Throwing rocks at bottles causes bottles to break". The relata of singular causal relations are events that actually occur (understood in a certain way, see below) rather than properties (such as being a smoker or breaking) that can stand in stand in general causal relations. The second point is how events are understood. For the purposes of the theory, they are understood simply as the exemplification of some property or properties at some specified place and time (or interval). Thus, for our purposes, if x stands for a time/place pair (call it $\langle t_x, s_x \rangle$) and X stands for a property (which may be complex, consisting of conjuncts $X_1, X_2, \ldots$), then an event is specifiable by saying that X is exemplified at x; this will be symbolized as "Xx" in what follows. An event is understood simply as the exemplification of a property or properties at such and such a time and place.

Third, the probability trajectory theory is a "probability-increase" theory of singular causation (causes raise the probability of their effects), but not in the ususal sense in which probability increase is understood. In the usual sense, "C increases the probability of E" is understood in terms of conditional probability comparison: $\Pr(E/C) > \Pr(E/{\sim}C)$ (or, equivalently, $\Pr(C/E) > \Pr(E)$). With suitable qualifications (see below), I think this understanding of probability increase is suitable for understanding probabilistic causation at the type level, that is for understanding claims of the form, for example, "(property) C is a positive causal factor for (property) E". For the token level (for singular causation), however, the probability trajectory theory focuses on the actual evolution of the probability of the token effect from around the time of a candidate token cause event to the time of the token effect event. A probability trajectory is the shape of the time/probability graph that represents such an evolution.

At this point, it will be useful to point out that probabilistic claims must be understood as relative to a population, P, understood to be of a certain kind, Q – where it is the kind Q that really controls, so reference to P will be suppressed in what follows and I will refer simply to "populations Q".[3] For example, certainly the probabilistic (as well as the type level causal) relations between smoking and heart attacks are different in the human (Q) population from what they are in a population of smoking machines (Q') in a laboratory. Thus, type level causal claims should be taken to be of the form "C has such and such a kind (positive, negative, etc., see below) of causal significance for E in population Q" (where the relevant probabilities are understood as relative to Q). At the token level, causal claims will be formulated as "*Xx* has such and such a kind of token causal significance for *Yy*", where the relevant population is the unit set $\{\langle x,\ y\rangle\}$ whose kind can often be taken to be understood by context (as in type level claims) in a request for causal information (in a question of the form "What was the causal role of *Xx* for *Yy*?"). Then, with *X, x*, and *y* left implicit, I let "$\Pr_t(Y)$" symbolize the probability, at time t, that the second member, y, of $\langle x,\ y\rangle$ in the singleton population $\{\langle x,\ y\rangle\}$ exhibits (or will exhibit) property Y (the role of X and x in this will be explained shortly). $\Pr_t(Y)$ can be thought of as a conditional probability, $\Pr(Y/K_t)$, where K_t is the conjunction of relevant (to Y) factors that have fallen into place by time t – but there are two further important features of $\Pr_t(Y)$, to which I now turn.

The fourth point about the probability trajectory theory I need to mention is two qualifications analogous to those alluded to above in connection with the type level probabilistic theory of causality. In the type level theory, it will not do simply to say that C is a positive causal factor for E in population Q if and only if $\Pr(E/C) > \Pr(E/{\sim}C)$, for there is the possibility of "spurious correlation". E can be positively probabilistically correlated with C even when C is not a cause of E, for example if there is a common cause of C and E (for example, rain is positively correlated with falling barometers not because the latter causes the former, or vice versa, but because the two have a common cause, an approaching cold front, or falling barometric pressure). The standard solution to this problem of spurious correlation is to "hold fixed" (conditionalize on) the appropriate factors when making the relevant probability comparisons. Without going into detail, I simply describe the adjustment (for the type level theory, preliminary to the token level theory) to the "basic probability increase idea" (causation goes by correlation) in three steps.[4] First, let $F_1, F_2, \ldots, F_n$ be those factors that need to be "held fixed" (what they are is given in the third step below), and let $K_1, \ldots, K_m$ be the maximal conjunctions of the F_i's and their negations that have positive probability both in conjunction with C and in conjunction with ${\sim}C$ (there are $m < 2^n$ K_j's). The K_j's are called causal background contexts. Second, C is called a "positive", "negative", or "neutral causal factor" for E if and only if, for all j, $\Pr(E/Kj\&C)\ >, <$, or $= \Pr(E/K_j\&{\sim}C)$. The feature of this according to which the inequality or equality must hold across *all* causal background contexts is called "*context unanimity*". In case unanimity fails, C is called "causally mixed" for E. And third, we must say what factors should be included among the F_i's. These are all factors that are both (i) causally independent of C (C is causally neutral for them) and also (ii) either (iia) causally relevant (positive, negative, or mixed) for Y or (iib) interactive for Y with respect to X. A factor F is interactive for Y with respect to X if the comparison (greater than, less than, equal to) between $\Pr(E/F\&C)$ and $\Pr(E/F\&{\sim}C)$ is different from what it is between $\Pr(E/{\sim}F\&C)$ and $\Pr(E/{\sim}F\&{\sim}C)$ (i.e., the comparison between $\Pr(E/C)$ and $\Pr(E/{\sim}C)$ is different in the presence of F from what it is in the absence of F).[5]

As to (i) just above, we should not hold fixed factors causally intermediate between C and E in evaluating the causal role of C for E; otherwise, we "rob" C of the probabilistic impact on E that it should have in virtue of its causal impact, if any, on E. (iia) and (iib) just above are the two qualifications alluded to above for the type level theory. (iia) is intended to handle the usual kinds of spurious causation, and (iib) is intended to get the theory to give the right answer of "mixed" causal factorhood when this is the truth.

For our purposes here, the important point (this fourth point about the *trajectory* theory for probabilistic causation on the *token* level) is that qualifications analogous to (iia) and (iib) – subject still to (i) – apply also at the token level. That is, in assessing the token causal significance of an event *Xx* for an event *Yy*, we must control for events that occur causally independently of *Xx* and that bear possibly confounding casual relations (relative to *Xx*) for *Yy*. The best way to implement these requirements, I think, is to hold fixed (positively or negatively, depending on how things actually happen) factors Z such that *Zz* (for the relevant time and place *z*) actually occurs *token causally independently* of *Xx* and the factor Z is either (again) either (iia) a positive, negative or mixed cause of *Y at the type level* or (iib) interactive for Y with respect to *X at the type level*. Of course, at the token level, there is just one relevant causal background context, call it K_a, which corresponds to features of the way things actually are in the actual situation in question. Then the relevant probability trajectory traces the evolution of $\Pr(Y/K_a \& K_t)$ as t varies from around the time of *Xx* to the time of *Yy*). As in the type level theory, the background context is a very important feature of the token level theory, and the relevant qualifications on the basic probability increase idea are natural given their analogs in the type level theory.[6]

Fifth, and finally, the taxonomy of kinds of causal significance is somewhat different at the token level, according to the trajectory theory, from the way it is in the type level theory. This is because of the different conception of probability change used at the token level. At the type level, there are, qualitatively, four kinds of probabilistic impact that C can have on E: unanimously positive, unanimously negative, unanimously neutral, or nonunanimous (for positive, negative, neutral, and mixed casual factor-hood, respectively, and of course there are several ways in which C can be nonunanimous for E, both probabilistically and causally). These are different kinds of conditional probability comparisons across contexts. The token level theory, on the other hand, pays attention to the way the probability of a later event *Yy* actually evolves from around the time of an earlier event *Xx* to the time of *Yy*. And qualitatively speaking, there are four basic shapes such a probability trajectory can assume: (1) it can be higher just after the time of *Xx* than it was just before that time and stay higher all the way until the time of *Yy*, (2) it can be lower just after the time of *Xx* than it was just before that time, (3) it can be the same just after the time of *Xx* as it was just before that time, and (4) it can be higher just after the time of *Xx* than it was just before that time but not remain higher than that previous value all the way until the time of *Yy*. At the time t of *Yy*, the probability of Y becomes 1 ($\Pr(Y/K_a \& K_t) = 1$ when t is the time of *Yy* or after that time). Note also that value of the probability of Y at the time of *Xx* does not enter into the theory; this I take to be equal to $\Pr(Y/K_a \& K_t)$, where t is the time of Xx, and K_t includes X but does not register the (perhaps improbable) actual consequences of the occurrence of X at that time. I have called (just to pick some suggestive terminology) the four kinds of causal significance corresponding to (1)–(4) above Yy's happening *because of, despite, independently of*, and *autonomously of* Xx, respectively. These kinds of causal significance can come in degrees, but I will not enter into that here except to say that the degrees can be measured, basically, by the magnitudes of absolute probability differences for the candidate effect event across the time of the candidate cause event.[7]

I should point out that the explications given above of the various token and type level causal concepts are not intended to be definitions – as such they would be circular, of course. Rather, they should be understood as constraints on the relationships among probabilistic and causal relationships. And finally, I point out that the probability trajectory theory is supposed to apply only in cases of nondeterministic causation – for deterministic causation, the probability trajectories would be trivial and the differences between the four kinds of token causal significance described above could not show up.

Probabilistic Causal Preemption

For a long time, the phenomenon of preemption, described above, has provided test cases for theories of causation. Mackie (1974, pp. 44–45) describes several examples, and gives references dating back to the 1920's. And, as mentioned above, the problem of preemption has been used recently to challenge the probability trajectory theory just outlined. I believe that the probability trajectory idea has the resources to deal with the phenomenon – to give the right, indeed the intuitively right, answers about what causes what, when the questions and answers are properly formulated and understood. I begin by giving three examples described by Mackie, and then turn to a couple of more recent examples. The numbering below follows Mackie (1974, p. 44).[8]

(iii) '… conditions (perhaps unusual excitement plus constitutional inadequacies) [are] present at 4.0 p.m. that guarantee a stroke at 4:55 p.m. and consequent death at 5.0 p.m.; but an entirely unrelated heart attack at 4.50 p.m. is still correctly called the cause of death, which, as it happens, does occur at 5.0 p.m.'

(iv) Smith and Jones commit a crime, but if they had not done so the head of the criminal organization would have sent other members to perform it in their stead, and so it would have been committed anyway.

(v) A man sets out on a trip across the desert. He has two enemies. One of them pours a deadly poison in his reserve can of drinking water. The other (not knowing this) makes a hole in the bottom of the can. The poisoned water all leaks out before the traveller needs to resort to this reserve can; the traveller dies of thirst.

Let us focus on case (v) for now. In this case it is supposed to be clear that it is the puncturing of the can, not the poisoning of the water in it, that caused the death, though either alone (or as in the example both together) would have sufficed. Let us see what the probability trajectory theory has to say about this. And for this purpose let us assume that all the causal relations involved are probabilistic and that all the relevant probabilities are nonextreme (not 0 or 1) – that is, until the time of an event, at which time its probability assumes the value 1. First, the theory correctly rules that the poisoning of the water in the can did not cause the death. This is because the puncturing of the can is a cause of death that is causally independent of (not an effect of) the poisoning (the second enemy didn't even know about the poisoning). In fact, all of the relevant effects of puncturing the can, including the presence of terminal dehydration, are causally independent of the poisoning and also causally relevant to the death. Thus, all these factors (including the puncturing and the dehydration) have to be held fixed (conditionalized on) when assessing the probability-trajectory probabilities relevant to assessing the causal role of the poisoning for the death. And given the presence of all these factors, the poisoning does not change the probability of death across the time of the poisoning. So the theory correctly rules that the death is causally independent of the poisoning of the water in the can.

What about the token causal role of the puncturing of the can for death, according to the probability trajectory theory? In this case, we have to hold fixed the poisoning of the water in the can, for this is causally relevant to death and causally independent of (not caused by) the puncturing. There are two things to say here. *First*, focusing on the causal role of the puncturing for the *factor of death*, if the poisoning does not necessitate death, then there is still room for the puncturing to increase (or to decrease, see below) the probability of death, even conditional on the poisoning; but if the poisoning is highly efficacious in producing death (at the type level), then any increase in the probability of death across the time of the puncturing would be very small, but the theory would still rule that the death is because of the puncturing, but only to a small degree. Further, however, there is the possibility that, if the poison is more efficacious in producing death than the puncturing of the can is, then the probability of death could actually *decrease* across the time of the puncturing (since then the man becomes no longer vulnerable to the poison but only to the less efficacious cause of death, the puncturing of the can). In that version of the example, the trajectory theory would rule that the death is *despite* (to some small degree I suppose) the puncturing. And if the poisoning and puncturing are equally efficacious for death, then the probability of death could remain the same across the time of the puncturing and the theory will say that the death is causally independent of the puncturing. I think all this (the verdicts of the trajectory theory in the various versions of the example) is correct, when we focus simply on the factor of *death* as what was exemplified by the man at the relevant later time and place. But *second*, if this seems unintuitive (that the degree of causal significance of the puncturing should be called "small" in the example, or that the death should be called even a little "despite", or even "causally independent of", the puncturing), then I think it does so only because it leaves out the rest of the causal story as seen from the point of view of the probability trajectory theory. There are of course many factors that are exemplified at the relevant time and place, including not only death but also the *factor of death-accompanied-by-dehydration*.[9] And of course, even holding fixed the factor of poisoning, we should expect the probability of death-accompanied-by-dehydration to rise considerably across the time of the puncturing, so that the probability trajectory theory will give the correct answer that the death-by-dehydration was, to a significant degree, "because of" the puncturing.

Thus, in example (v), the trajectory theory gives the clearly correct answer about the casual role of the poisoning for death, and when the relevant factors are isolated and the relevant questions asked, the theory gives the correct answers about the causal role of the puncturing for death, and for death-accompanied-by-dehydration. Examples (iii) and (iv) can be handled in analogous ways, when understood to involve nondeterministic causation and nonextreme probabilities. In (iii), the death is, intuitively and according to the theory, casually independent of those conditions that were right for a stroke (hold fixed the "entirely unrelated" heart attack and the resulting lack of blood circulation that eventually and more proximately led to the death); and the death is, to some small degree, either because of or despite (or even independent of, depending on the details of the example) the heart attack (depending on the relative efficacies of the heart attack and the pre-stroke conditions for death); but death-accompanied-by-lack-of-blood-circulation is, to a high degree, because of the heart attack (hold fixed, of course, the pre-stroke conditions). In (iv), the crime is, intuitively and according to the theory, causally independent of the backup plans of the head of the criminal organization (hold fixed Smith and Jones' intentions and the successes in the various steps along the way that culminated eventually in the crime); and the mere fact of the crime is, to some small degree, either because of or despite (or even independent of, depending on details of the example) Smith and Jones' forming the intention to commit it (depending on the relative skills of Smith and Jones compared to the backup crew); but the crime, in the exact way it was committed by Smith and Jones, is, to a high degree, because of Smith and Jones' forming the intention to commit it (hold fixed, of course, the backup plans of the head of the crime organization). (If however, the plans of the head of the criminal organization in some way contributed to Smith and Jones' making their plans, then the story would be different and we would not hold fixed Smith and Jones' plans in assessing the causal role of boss' plan for the crime, and we would get the right answer that the crime was because of the boss' plans.)

Note that there are three question/answer pairs addressed in the above analysis of the three examples: (1) What is the causal role of the (preempted) event $X'x'$ (poisoning, conditions being right for a stroke, the boss' plans) for effect Yy (death, death, crime) (2) What is the causal role of the (preempting) event Xx (puncturing, heart attack, Smith and Jones' plans) for the effect Yy, and (3) What is the causal role of (preempting) event Xx for the effect event $Y'y$ considered in a more precisely specified way (death accompanied by dehydration, death accompanied by lack of blood circulation, crime in the specific way committed by Smith and Jones). The application of the trajectory theory to questions (1) went the smoothest (by holding fixed the preempting cause and its effects intermediate between it and the final effect). The application to questions (2) was fairly straightforward as well, except that it initially seemed that the trajectory theory was not giving the preempting cause Xx its due, in not assigning it a strong enough causal role in the production of the effect event Yy. However, answers to questions (3) were supposed to fix this seeming lack of match with intuitions by pointing to a more detailed, or different, specification of the way things were at time/place y for which Xx really *was* strongly causally responsible: the idea is that in fact Xx was *not* strongly causally responsible for y's being Y but *was* strongly causally responsible for y's being Y'. Applications of the theory to questions/answers (1) had to do with verifying that the theory does not say that *the preempted cause is a cause*; and applications of the theory to questions/answers (2) and (3) had to do with verifying that *the theory does say that the preempting cause is a cause (of the effect event appropriately understood).*

While issue (1) seems to be handled just fine by the theory, and the same for (2) given the assumed probabilistic nature of the examples, the application of the theory to issue (3) relies on a substantive assumption, which may be construed either metaphysically or empirically, depending on how one wants to construe the theory. That assumption may be formulated like this:

Trace Assumption In cases in which Xx preempts $X'x'$ in the production of Yy, there is some feature Y' of what happens at y that physically traces back to Xx and not to $X'x'$ and would not have been present at y had Xx not occurred and $X'x'$ caused Yy instead.

This formulation involves the ideas of "physically tracing back" and "counterfactual dependence"; these ideas are not ingredients in the trajectory theory, but rather this formulation is simply intended to use some (somewhat vague) ideas that are ingredients in other theories of causation and in our ordinary concept of causation and in terms of which we can test the implications of the trajectory theory. In the can of water example, (v), Y' was death specifically accompanied by dehydration; in the patient example, (iii), Y' was death specifically accompanied by lack of blood (ordinarily supplied where and in the manner the heart supplies it); and in the crime example, (iv), Y' had to do with some supposed specific way in which Smith and Jones committed the crime that differs from the way it would have been committed if the backup plan had had to be implemented.

There are three points I would like to make in clarifying and defending this assumption. First, that factor Y' is *not* intended to involve a possible difference in time or place in which Yy did occur as a result of Xx and that would have been different had Xx not occurred and $X'x'$ been the cause of some $Y''y'$ instead. Throughout, I am assuming that we are concerned with the causal significance of what happened at one specific time/place, x, for what happened at another, y. And in fact, if, in the examples, the difference between Xx and $X'x'$'s being the cause did make a difference in the time at which an effect in question occurred, then I think the probability trajectory analysis would have an even easier time dealing with the relevant examples. As explained above, I am working with a conception of events on which they are individuated by a specification of a time/place (point or interval) and a set of factors exemplified at that time/place (like a trope): if either the time/place differed or the set of properties differed (even in one's being a subset of another) in two specifications, then we have specifications of two different events.[10] So if, for example, in the crime case, (iv) above, "the" crime would have been committed at a later time than Smith and Jones' crime had Smith and Jones failed and the backup plan been used, then there would be this temporal feature of the actual crime that traced back to Smith and Jones' plan but not to the boss' backup plan,[11] and surely that there even was a crime at the time/place y of the actual crime is a feature of y that traces back to Smith and Jones' intentions and that would not have been present at y were it not for their plans.

Second, it might be objected that in many cases of preemption there are traces of the preempted event that are present in the effect event that would not be there had the preempted event not occurred, so that my application of the trace assumption does not distinguish the preempted from the preempting event. So, for example, in the patient case, (iii) above, there may be features Y'' present in the patient at the time of his death that physically trace back to the (preempted) conditions-just-right-for-a-stroke-at-4:55-p.m.-and-death-at-5:00-p.m. present at 4:00 p.m., where Y'' would not have been present at y had these 4:00 p.m. conditions not been present. This seems natural enough, of course. And it is also natural enough to say that $Y''y$ is because of $X'x'$ but not because of Xx, while $Y'y$ (Y' again = death with lack of blood in the places where the heart ordinarily supplies it in the way it ordinarily does) is because of Xx but not because of $X'x'$, which is just what the probability trajectory theory yields.

And third, we may wonder whether the trace assumption is always true, whether there really could not be cases in which *everything* that happens at y is just the way it would have been had Xx not occurred and $X'x'$ had been the cause of y's being Y and y's having all the other features Y' it actually has. I know of four examples that have been described recently with this issue specifically in mind. I will consider two of these here.[12]

Paul (2000) describes an example involving two cats and a fly:

> C. Louise crouches aiming for [a] fly. Possum also crouches, aiming for the same fly. C. Louise jumps. Possum, who has been practicing, jumps a moment later, but his (newly acquired) agility makes him able to catch the fly at the same time as C. Louise. Unfortunately for Possum, there is a little known law that states that flies, when pounced upon by multiple cats, are captured by the cat who jumps first. Since C. Louise jumps before Possum, she gets the fly. If C. Louise had not jumped, Possum would have captured the fly in the very same way and at the very same time. C. Louise's pounce, albeit through no intrinsic feline merit, trumps Possum's. (Paul 2000, p. 247)

But surely there are features of the actual fly-catching that trace back to C. Louise's pounce and not to Possum's and that would not have been present had C. Louise not pounced and Possum had caught the fly. For example, it was C. Louise's paw (or teeth, or however the cat did it) that actually made contact with the fly; this clearly traces back to C. Louise's pounce, not Possum's, and would not have been present had C. Louise not pounced and Possum had caught the fly instead.[13]

Ned Hall (2004) says, "it's easy enough to construct cases in which **c** is clearly a cause of **e** but in which neither **c** nor any event causally intermediate between it and **e** make the slightest difference to the way **e** occurs", and he attributes to Steve Yablo the following modification of the story of Billy and Suzy (described at the beginning of this paper):

> This time Billy throws a Smart Rock, equipped with an on-board computer, exquisitely designed sensors, a lightning-fast propulsion system – and instructions to make sure that the bottle shatters in exactly the way it does, at exactly the time it does. In fact, the Smart Rock

> doesn't need to intervene, since Suzy's throw is just right. But had it been any different – indeed, had her rock's trajectory differed in the slightest, at any point – the Smart Rock would have swooped in to make sure the job was done properly. (Hall 2004)

But again, I think that when we think carefully about what happened at y (the time and place of the bottle's shattering) we will find there features that trace back to Xx (Suzy's throwing her rock) that would not have been present had she not thrown her rock, or if for some other reason the Smart Rock had done the job instead. For example, Suzy's rock is in front of the Smart Rock – closer to the bottle than the Smart Rock is just before the shattering, and in contact with the bottle, while the Smart Rock is not, at the time of the initial rock-bottle contact. And just after the initial rock-bottle contact, surely the configuration of Suzy's rock, the Smart Rock, and the shards of glass is different from the way it would have been had Suzy not thrown her rock, or if for some other reason the Smart Rock had done the job instead. So, realistically and carefully understood, it seems that there are features Y' just before, at the time of, and just after the shattering that trace back to Suzy's throw and that would not have been present had Suzy not thrown her rock, or if for some other reason the Smart Rock had done the job instead. And the probability of these features Y' of y would seem to increase across the time at which Suzy actually succeeds in letting her rock fly. In Section 3, however, I will discuss the theoretical possibility of examples in which there are no such factors Y', that trace back to Xx and that would not have been present were Xx not to have occurred.

The discussion here may remind readers of David Lewis' discussion of *fragility* of events (1986, p. 196ff.). An event is "*fragile* if, or to the extent that, it could not have occurred at a different time, or in a different manner. A fragile event has a rich essence; it has stringent conditions of occurrence" (p. 196). The idea I think is that what happens at a time and place y can be considered in a more or less detailed way, Y, so that at that same time and place both more and less fragile events occur. ("Don't say: here we have the *events* – how fragile are *they*. ... Properly posed, the question need not have a fully determinate answer, settled once and for all. Our standards of fragility might be both vague and shifty (pp. 196–197)). And Lewis considers a strategy like the one above involving our trace assumption, and asks, "Wouldn't we still have residual cases of redundancy [of which preemption is one kind], in which it makes *absolutely* no difference to the effect whether both of the redundant causes occur or only one?", and answers, "Maybe so; but probably those residual cases would be mere possibilities, far-fetched and contrary to the ways of this world" (p. 197, Lewis' italics). The trace assumption is that such far-fetched cases just do not happen (but, again, I will consider such theoretically possible cases in Section 3). But Lewis goes on to say that the strategy makes for more trouble than it cures anyway, and he considers two examples intended to illustrate this, to which I now briefly turn.

In the first example, one gentle soldier in an eight-soldier firing squad did not shoot. If the victim's death is considered to be a very fragile event, then a seven-bullets-through-the-heart death is a different event from an eight-bullets-through-the-heart death. "So the gentle soldier caused the death by not shooting, quite as much as you caused it by shooting! This is a *reductio*" (p. 198). In the other example, Boddie eats first a large dinner and then poisoned chocolates. Boddie then dies from the poison, but the large dinner slowed the absorption of the poison in the chocolates and the death occurred somewhat later and in a slightly different manner than it would have without the large dinner. "If the death is extremely fragile, then one of its causes is the eating of the dinner. Not so" (p. 198). Leaving aside the factor of time (as before), the number of bullets through the heart and the exact manner of Boddie's death are factors Y' that trace back to the gentle soldier's not shooting and the eating the large dinner, respectively, and which would not have been exemplified at y (times/places of the deaths) had the failure to shoot or the eating of the large dinner not occurred.

In these two examples, as in the other examples considered above, there are various things going on in, or features of, the spatio-temporal region y of the effect events, Yy, $Y'y$, and so on, as well as various things going on in, or features of, the spatio-temporal region x of the cause events, Xx, $X'x$, and so on. And different features X of the earlier time/place x have different causal significances for different features Y of the later time/place y. While it does indeed sound unusual to say that the gentle soldier's not shooting caused *the victim's death*, or that Boddie's large dinner caused *Boddie's death*, I suggest that this is only because the italicized phrases here single out or bring to mind a feature of the later events – namely $Y = death$ – for which the earlier event was not causally responsible. But, as in the previous examples, there are also features – $Y' =$ death-with-exactly-seven-bullets-through-a-heart and $Y' =$ death-by-poison-mixed-with-that-large-dinner – for which the cause events Xx were responsible. I find it completely natural to say that the gentle soldier's omission was responsible for there being only seven bullets penetrating the victim's heart and that Boddie's large dinner caused his death in the way that the death in fact occurred – all under the circumstances (the other seven soldiers' shooting and Boddie about to eat the poisoned chocolate). I agree that fragility is often a matter of vagueness or shiftiness, but when the relevant features of a cause event and an effect event are settled either by context (and the relevant features can shift from context to context) or by asking a precise question (of the form "What was the causal significance of x's being X for y's being Y?"), then a precise answer (which specifies each of x, X, y, and Y) should not, I think, sound unnatural at all.[14]

Before leaving the topic of ordinary preemption (or cases in which it seems, to me at least, that a plausible candidate for a Y' of y tracing back specifically to Xx can be found), I consider two recent examples specifically addressed to probabilistic causation and with the trajectory theory in mind. Christopher Hitchcock describes the following scenario (similar to the Billy and Suzy example described above but not exactly a case of preemption):

> Suppose that two gunmen are shooting at a Ming vase. Each one has a fifty percent chance of hitting the vase, and each one shoots independently, so the probability that the vase shatters is 0.75. (For simplicity, we will ignore the possibility that the vase might survive a bullet hit.) As it happens, the first gunman's shot hits the vase, but the second gunman misses. (2004)

In this case, in evaluating the causal role of either gunman's shooting for the shattering of the vase, we must hold fixed, for one thing, the other gunman's shooting, for the latter is causally independent of the former and causally relevant to the vase's shattering. This gives the right answer for the first gunman's shooting: the probability of the vase's shooting increases from 0.5 to 0.75 across the time of his shooting (and I suppose doesn't decrease between that time and the time of the vase's shattering). But at first it may seem that the same is true for the *second* gunman's shooting, and that the trajectory theory would give the wrong answer that the vase's shattering was because of *this* shooting. (As before, I ignore the time element, assuming that the time of the shattering would be the same no matter which gunman's bullet hit the vase.)

However, there are two ways in which the trajectory theory blocks this conclusion. First, it is after all *the first gunman's bullet* that hits the vase; this is a feature of what happens at the place/time y of the shattering that traces back to the first gunman's shooting and that would not have been present at y had the first gunman not shot, and whose presence at y the theory says is because of first gunman's shooting, and of course *not* because of the second's. This is similar to what I said about the version of the story of Billy and Suzy with Billy's Smart Rock. And second, recall that we must hold fixed *all* factors whose exemplifications are token causally independent of a candidate cause and which are causally relevant to the relevant factor exemplified at the later time. In this case, the first gunman's bullet entering or being just about to enter the vase (just before the vase actually shatters) is such a factor. And conditional on this factor, the second gunman's shooting does not affect the probability of the vase's shattering across the time of his shooting, which makes the theory give the correct answer that the shattering was token causally independent of the second's gunman's shooting.[15] In addition, presumably at some time after the two shots, the bullet from the second gunman is off course; this should be held fixed in evaluating the causal role of the first gunman's shooting for the shattering, in which case the probability of the shattering increases from 0 to 0.5 at the time of the first gunman's shooting. This would seem to be the correct way to apply the theory to the question of the casual significance of the first gunman's shooting for the vase's shattering, in which case again the theory says that the shattering was because of the first gunman's shooting.

Douglas Ehring (1994, 1997) describes a somewhat more complicated example. A very sick patient in a hospital is connected to a mechanism that can deliver drugs to the patient. The mechanism is connected to two sources of drugs, a source of drug A and a source of drug B. If the patient receives neither of these drugs, then the chances of survival are very low; but if the patient gets either or both of the drugs, then the chances of survival are 0.5; and there is a law of biology according to which, given the patient's condition, the chances of survival cannot possibly exceed 0.5 at this time. It is improbable that both drugs would flow to the mechanism that is connected to the patient. And if both drugs are released from their sources, then there is a chance that the drug that reaches the mechanism first will set off a device that blocks the flow of the other drug to the patient. The release of either drug from its source is causally independent of whether the other drug is released from its source. And assume finally that if at least one of the drugs is released from its source then the probability is 1 that the patient will receive one of the two drugs through the mechanism (drug A if only A is released from its source, B if only B is released from its source, and at least one of the two if both are released from their sources). Here is what happens. First, a valve is opened that starts the flow of drug B to the mechanism; then (improbably) a valve is opened that starts the flow of drug A to the mechanism. But drug A reaches the mechanism first and this (improbably) blocks drug B from flowing through the mechanism to the patient; drug A reaches the patient and the patient survives. Ehring says (correctly, in a way) that the release of drug A is what actually saved the patient. However, at the time of A's release, the probability that the patient would survive was already at its maximum value of 0.5, since drug B was already on its way to the mechanism, guaranteeing that at least one of the two drugs would reach the patient. Thus, the release of drug A did not, at the time of its occurrence, alter the probability that the patient would survive. This is supposed to be a counterexample to the probability trajectory theory since the release of drug A caused the survival but the probability trajectory theory says that survival was causally independent of the release of A.

However, I think the right way to view this example is this: the release of drug A *did not* cause (exactly) *the survival*, but it did cause *the mechanism to deliver the drug it did (A)* to the patient, it did cause the mechanism to start delivering the drug it did *at the time it did (just before drug B reached the mechanism)*, it did cause *the mechanism to block the flow of drug B*, and it did cause the patient to *survive with drug A in his veins*. Even though the probability of survival cannot be increased above the stipulated biological limit of 0.5 by the release of drug A, the probabilities of these other factors can be increased by this from what their values were before the release of drug A, but of course not of the last other factor to above 0.5. Again, if we pay attention to the fact that many properties are exemplified at the relevant later time/place y (in this example, the spatio-temporal region within which the patient recovered), then, when we tell the whole story, the probability trajectory theory will deliver the correct answers, in the light of which some initially perhaps misleading answers (in this example, that the release of drug A *did not* in the circumstances cause, exactly, the patient to survive) are no longer misleading. (I leave it to the reader to ponder the application of the probability trajectory theory to the causal role of the release of drug B for the exemplifications of the various factors exemplified around the patient's recovery.)

I conclude that in cases in which the trace assumption holds, and when causal questions and answers are formulated precisely (including an x, an X, a y, and a Y), the trajectory theory gives the correct answers – indeed the *intuitively* correct answers in light of answers to other questions, that is, in light of the "whole story".[16]

Failure of Tracing, and Symmetric Overdetermination

In this section I discuss two matters that were alluded to above but postponed, namely the theoretical possibility of failure of the trace assumption and the phenomenon of symmetric overdetermination.

Theoretical Possibility of Failure of the Trace Assumption Suppose that, somehow, an effect e would have occurred in exactly the same way if, instead of the preempting cause c causing e, the preempted cause c' had caused e. Expressed in terms of the notation introduced above, the situation is this: for *all* factors Y present at time/place y, y would still have exemplified Y had $X'x'$ not been pre-empted and Xx was not causally responsible for what went on at y. As suggested in the discussion of examples above, I think this would be extremely rare in a case of (at least late) preemption,[17] so I will turn to a different kind of case, an example basically the same as one I discussed briefly in my (1991, pp. 384–386) that I think will serve our purposes here but which I am not sure should be counted as a case of preemption.[18]

Suppose a golf ball is rolling straight toward the cup with a 95% chance of falling in. A squirrel kicks the ball away, but improbably enough the ball comes off the squirrel's foot with the same 95% chance of falling into the cup, but along a different path (made possible by the contour of the golf course). Suppose further that this new 95% chance is for the ball's landing in the cup in exactly the same way ("locally") as it had a 95% chance for before the kick (i.e., crossing the same point on the rim of the cup, from the same direction, with the same speed, at the same time, etc.). The ball falls in the cup.[19] In one sense, we want to say that the squirrel kick caused the birdie, but on the other hand it is part of the example that the actual effect of the kick is to leave the probability of the birdie's occurring, in exactly the same way as it probably otherwise would have, unchanged. Even though the birdie's happening exactly the way it did traces back to the squirrel kick, the net actual effect of the kick is to leave the probability of this happening unchanged. Because of the latter, the trajectory theory says the birdie occurred (exactly as it did) "token causally independently" of the kick. This seems to be a correct verdict of the trajectory theory in one sense of cause: the probability of things being the way they actually turned out to be at the time of the birdie was left unchanged by the squirrel kick, given the way the ball was moving after the squirrel kick. But there is still the tracing-back intuition according to which the squirrel kick *is* relevant to the birdie, even though there is no feature of the effect that traces back to the squirrel kick and that would (with the same probability as conferred by the ball's original motion) not have been present had the squirrel not kicked the ball.

My intuition here is that, as far as causation strictly speaking in concerned, the squirrel kick *is* irrelevant to the birdie's happening, in the exact way that it did – that being, again, because the kick left the probability of the ball's falling into the cup, in exactly the way it did, unchanged. On the other hand, there is the "tracing-back" intuition that the squirrel kick is somehow relevant to the birdie, for the kick was responsible for the ball's taking the path it actually did take into the cup. If we want to explain just how or why the ball fell into the cup, we would surely want to include the fact of the kick and the fact of the path the ball actually took in moving to the cup. The way I have just expressed the way in which I think the kick is relevant to the birdie suggests my diagnosis of the case: we should separate the concepts of causation and explanation here. It would be a project in itself to develop this suggestion in detail. I suggest that we should separate question of the explanatory contribution of the kick for the birdie (in the exact way that it happened, in this unusual case) from the question of causal impact of the kick on the birdie (the kick's having no actual effect on the probability of the birdie in the exact way that it happened, in this unusual case).

I think this intuition and suggestion applies also to the theoretically possible kind of preemption in which the trace assumption fails. In such cases (if there are any), y's being Y (for all relevant Y's exemplified by y) traces back to and is partially *explained* by Xx, even though, due to the presence of the backup event $X'x, Xx$ did not, in these circumstances, positively *causally* contribute to y's being Y. I think that this tracing back, explanatory kind of significance of an earlier event for a later event can be captured by a trajectory-style analysis. It is not clear to me exactly how this should be carried out, but it seems that the basic idea should be to point to chains or networks of trajectory-style connections between an earlier event and a later event. For example, while the birdie was not exactly because of the squirrel kick, there are intermediate events that were because of the kick and that the birdie was because of (thus, "because of" is not in general transitive).[20] But again, this is a topic for a separate project.

Symmetric Overdetermination I turn now to the second kind of example of overdetermination, exemplified in examples (i) and (ii) in Mackie's list of five examples, (iii)–(v) of which were discussed above.

(i) A man is shot dead by a firing squad, at least two bullets entering his heart at once, either of which would have been immediately fatal.

(ii) Lightning strikes a barn in which straw is stored, and a tramp throws a burning cigarette butt into the straw at the same place and at the same time: the straw catches fire. (Mackie 1974, p. 44)

Mackie says, "In these cases even a detailed causal story fails to discriminate between the rival candidates for the role of cause. We cannot say that one rather than the other was necessary in the circumstances even for the effect *as it occurs*" (p. 47, Mackie's italics). Of course the concern in this paper is with probabilistic causation, so let us understand the examples as probabilistic (what the members of the firing squad did makes the victim's death very probable and the lightning and the tramp's cigarette but each make the fire very probable), so that the necessity referred to in the quote from Mackie is not exactly to the point in the way we will understand the examples. The salient point about these examples is indicated by Mackie's italicized phrase. Unlike Mackie's other examples (where we can say that the effect event was a heart-attack-death and not a stroke-death, the effect was a Smith-and-Jones-crime and not a back-up-crew-crime, and the effect was a dehydration-death and not a poison-death), examples (i) and (ii) are supposed to be cases in which the effect cannot be traced back to one of the two candidate causes and not the other.

I think the same considerations applied above to cases of preemption, or asymmetric overdetermination, can be applied also to examples like (i) and (ii), and that the probability trajectory theory, properly applied, yields intuitively correct verdicts. First, although it is not specified in the examples, it seems natural to think that, understanding the examples probabilistically, the combination of two causes makes for a higher probability of the effect than one of the two alone does (of course it is natural for this unspecified possible feature of the examples not be specified by Mackie, since he was operating with a necessity/sufficiency conception of causation). Also, in evaluating the causal role of one of the earlier events for the later one we should hold fixed the other earlier event (since it is at least implicitly assumed in the examples that the two earlier events are causally independent of each other). In this case, we should say that all of the earlier events were causes of the respective later events in their examples. And the probability trajectory theory applied to these examples says that the victim's death was because of each of the bullets' entering the heart (each raised the probability of death over what it is conditional only on the other) and the fire was because of the lightning and also because of the cigarette butt (each raised the probability of the fire over what it is conditional only on the other). However, the examples could be understood in a way that, given that one of the earlier events occurs, the other cannot further increase the probability of the relevant later event (as in Ehring's example in which there is no possibility of raising the probability of the patient's survival above 0.5). Also, as in previous examples, one could naturally complain that this approach does not assign an intuitively high enough causal significance to the causes for the effect.

But second, we could try applying the trace assumption to these examples. Mackie says that the two candidates for a cause cannot be discriminated between even by considering the effect "*as it occurs*", by which I take him to mean, "paying attention to all of the features of the later event". Of course it is worth pondering this theoretical possibility, but again I find it hard to imagine that the effect event would be exactly the same, in all its details, if either one of the two candidate causes occurred without the other. Surely there is some feature Y' of the firing squad's victim's death that would have been different had only one bullet pierced his heart (e.g., a difference in the change of momentum undergone by a part of the victim's body) and that traces back to a second bullet. And in the fire example, surely the presence of the lit end of the cigarette butt (or of the lightning) made for a difference Y' in what things were like at the time and place y of the fire's ignition that would not have been present if only the lightning (or lit end of the cigarette butt) had been there to start the fire, and which traces back to the presence of the cigarette butt (the lightning strike). And these differences Y' can be expected to reverberate into the exact future of the victim's body and into the exact ensuing course of the fire. In this case, we can naturally say – and the probability trajectory theory applied to the examples yields these verdicts – that the feature Y, the firing squad's victim's being dead, of the time and place of what happened to the victim is token causally independent of each of the bullet's entering the heart (under the circumstances of the other bullet's entering the heart), while the feature Y', above, of this time and place was because of a given bullet's entering the heart, and similarly for the fire example.

But third, we should not ignore the theoretical possibility of failure of the trace assumption in examples like this. In this case, I would again, as above, appeal to a distinction between a purely causal concept and an explanatory concept. If it is really true that neither of the "candidates for the role of cause" made any difference at all in the effect event (under these circumstances of the presence of the other candidate) then it seems, at least to me, reasonable to say that neither was a *cause* (under the circumstances), though a full *explanation* of just how and why the victim died or of how and why the fire was started would include both "candidates".

Conclusion

The propensity trajectory theory of singular causation provides a natural framework which supports intuitive judgments concerning causal roles of preempting, preempted, and symmetrically overdetermining evens. (And the phenomenon of preemption poses no special problem for the propensity trajectory theory of singular probabilistic causation.) Some of the conclusions above depend on the trace assumption, and given the theoretical possibility of failure of the trace assumption, on a suggested way of separating purely causal from otherwise explanatory roles of events for events.

Acknowledgements I thank Dan Hausman, Christopher Hitchcock, Stephen Leeds, Alfred Mele, and Elliott Sober for useful comments or discussion.

Notes

[1] I note here, for reference in a later note, that preemption in a case like this is sometimes called "late preemption", in contrast to cases of "early preemption" in which a causal chain from the preempted event is "cut" by some effect of the preempting event at some time intermediate between the times of the preempted event and the effect, where had this "cutting" not occurred the preempted event would have succeeded in causing the effect in question. (See Lewis (1986), and also Lewis (2000) where he prefers the terminology "early and late cutting".) In the Billy and Suzy example, Suzy's rock did not interfere (at least "relevantly much" we might say) with the trajectory of Billy's rock.

[2] For details, see Eells (1991), chapter 6.

[3] For details, see Eells (1991), chapter 1.

[4] For details, and defense of the requirements, see Eells (1991), chapters 2 and 3.

[5] A more general idea of interaction is that the pairs $\langle \Pr(E/F\&C),\ \Pr(E/F\&{\sim}C)\rangle$ and $\langle \Pr(E/{\sim}F\&C),\ \Pr(E/{\sim}F\&{\sim}C)\rangle$ are different. More general still would be to define interaction in terms of partitions of factors F.

[6] See Eells (1991) for examples, specifically at the token level, that demonstrate the need for holding fixed independent causes and interactive factors.

[7] This will do for present purposes. Again, for details, see Eells (1991), chapter 6.

[8] Quoted from Mackie (1974, p. 44). These examples (iii)–(v) are what today we call "preemption", or "asymmetric overdetermination", where it is clear which of two earlier events is the cause of the later event. His examples (i) and (ii), to be considered in Section 3, are cases of what we now call "symmetric overdetermination", where we are supposed to have no definite intuitions about which of the earlier events is the cause. This terminology is due to Lewis (1986). Mackie quotes example (iii) from Scriven (1964), (iv) is from Marc-Wogau (1962), and example (v) is based on a modification by Hart and Honor (1959) of an example of McLaughlin (1925–1926).

[9] Mackie (1976, p. 46) makes much the same point when he distinguishes between the "facts", "[that] *he died*, and that *he died of thirst*".

[10] Compare Kim (e.g., 1973), though he individuates events by triples consisting of an individual (or individuals), a property (or properties), and a time.

[11] See Paul (1998, 2000) for an application of this feature of at least some cases of preemption to the counterfactual analysis of causation.

[12] The other two, by Schaffer (2000), involve either an admittedly unrealistic but logically possible situation involving magic and action at a distance or a possible world with a physics different from actual physics. While it is interesting to apply our concept of cause to such conceptually possible scenarios, I will restrict my attention here to what seem to be attempts to describe more realistic such situations. The first example to be described would seem to be a more realistic example of the kind that Schaffer has in mind and is intended to illustrate the possibility of "trumping", as Schaffer calls it.

[13] Paul points out that this example involves action at a distance and suggests that a different analysis is necessary for such causation than for causation that needs chains of events. In a note below, I will describe a version of the C. Louise/Possum case in which I think the trace assumption fails and there is no action at a distance (but which is not a case of *late* preemption).

[14] Compare Eells (1991, pp. 286–289).

[15] Hitchcock in fact makes this point for what he calls "ebb and flow in the probability pool" approaches to singular probabilistic causation, of which the probability trajectory theory is an instance.

[16] I note also the uniformity of this treatment of preemption across cases of "early" and "late" preemption. The versions of the Billy and Suzy story discussed are clear cases of late preemption, the case of the thirsty traveller and Ehring's patient example are clearly cases of early preemption, and in some of the other examples discussed, different verdicts concerning classification into late or early preemption may depend on different understandings or specifications of the details of the examples.

[17] But consider this modification of the C. Louise/Possum case of Paul's described above. Suppose C. Louise sees Possum on the way to the fly and pushes Possum away and is able then to accelerate so that the time of the effect would be the same no matter which cat got the fly. Further, more saliently, suppose the effect is not exactly the catching of the fly but rather the fly being crushed inside of a box by a mechanism that is activated outside the box by a button, where the exact way the fly is crushed is not affected by who or how the button is pushed. Of course this is not a case of *late* preemption (which was Paul's topic) but it would seem to be a case of failure of the trace assumption (and in which there is no action at a distance). So I do not deny the possibility of failure of the trace assumption in cases of (at least *early*) preemption.

[18] I will note below why I think it is unclear that this should be counted as a case of preemption.

[19] Why isn't this clearly a case of preemption, of the golfer's swing or of the motion of the ball just before the squirrel's kick, by the squirrel kick? Each of these candidates for the preempted cause seems clearly part of the actual cause of the birdie: on a natural understanding of the example, had the golfer not swung, or the ball not been moving in the way it was just before the kick, the action of the squirrel would not have resulted in a birdie. So the influences of these candidates for a preempted cause are clearly not totally preempted by the action of the squirrel.

[20] Hall (2004) distinguishes between two concepts of cause which he calls "dependence" (which is just counterfactual dependence) and "production" (which is a tracing-back idea). He gives (deterministic) examples of dependence without production; the example just discussed can be thought of as a (probabilistic) case of production without dependence.

References

Eells E (1991) Probabilistic causality. Cambridge University Press, Cambridge

Ehring D (1994) Preemption and Eells on token causation. Philos Stud 74:39–50

Ehring D (1997) Causation and persistence: a theory of causation. Oxford University Press, New York and Oxford

Hall N (2004) Two concepts of causation. In: Collins J, Hall N, Paul L (eds) Causation and counterfactuals. MIT, Cambridge, MA and London 225–276

Hart HLA, Honor AM (1959) Causation in the law. Oxford University Press, Oxford

Hitchcock C (2004) Do all and only causes raise the probabilities of effects. In: Collins J, Hall N, Paul L (eds) Causation and counterfactuals. MIT, Cambridge, Massachusetts and London, England 403–418

Kim J (1973) Causation, nomic subsumption, and the concept of event. J Philos LXX(8):217–236

Lewis D (1973) Causation. Reprinted in his: Philosophical papers, vol II, 1986. Oxford University Press, New York and Oxford

Lewis D (1986) Postscripts to “causation”. In his: Philosophical papers, vol II. Oxford University Press, New York and Oxford

Lewis D (2000) Causation as influence. J Philos XCVII(4):82–197

Mackie JL (1974) The cement of the Universe: a study of causation. Oxford University Press, Oxford

Marc-Wogau K (1962) On historical explanation. Theoria xxviii:213–233

Menzies P (1996) Probabilistic causation and the pre-emption problem. Mind 105(5):85–117

McLaughlin JA (1925–1926) Proximate cause. Harvard Law Rev xxxix, 149ff

Paul LA (1998) Keeping track of time: emending the counterfactual analysis of causation. Analysis LVIII(3):191–198

Paul LA (2000) Aspect causation. J Philos XCVII(4):235–256

Schaffer J (2000) Trumping preemption. J Philos XCVII(4):165–181

Scriven M (1964) Review of the structure of Science (ed: Nagel’ E). Rev Metaphys xviii, 403–424

Miraculous Consilience of Quantum Mechanics*

Malcolm R. Forster

Baby Epistemology

Think back to a time when mirrors were a new experience to us as children. How did we learn that the world we saw in the mirror was not a different world from ours, but a reflection of the world we already knew? How did we acquire the parsimonious view that reflections provided independent views of the same objects?[1] Presumably, it has something to do with the way that mirror images are correlated with our more direct perception of the objects.

Reichenbach (1938) put forward a similar idea when he imagined an observer enclosed in the corner of a cubical world, where objects in the external world cast shadows on the walls of the enclosure (Fig. 1). It seems plausible that much more information about external world would be available if external objects cast shadows on two walls of the enclosure, rather than a single shadow, or even copies of a single shadow. It seems that the inferential engine inside our heads tends to favor the judgment that they are the shadows of a single object, rather than shadows of different objects.

There is no good formal theory of how such inferences work, although they seem to conform to informal principles such as the principle of parsimony, or Occam's razor, which is usually taken to state that "Entities are not to be multiplied beyond necessity."[2] The rule is vague in crucial ways. For example, what counts as an entity? And under what conditions does it become necessary to multiply an entity?

Consider the first question more carefully. Do probabilities count as entities? Is there, in other words, a principle that says that probabilities are not to be multiplied beyond necessity? In Section on Consilience in Probabilistic Examples, I argue that such a principle is useful in causal modeling for it provides an account of the time asymmetry of cause and effect that does not rely on the concept of

M.R. Forster (✉)
Department of Philosophy, University of Wisconsin-Madison, Milwaukee, WI, USA
e-mail: mforster@wisc.edu

*This paper was written, in part, under a grant from the Graduate School of the University of Wisconsin-Madison, which I gratefully acknowledge. I would also like to thank Elliott Sober for helpful comments on an earlier draft.

E. Eells and J.H. Fetzer (eds.), *The Place of Probability in Science*, Boston Studies in the Philosophy of Science 284, DOI 10.1007/978-90-481-3615-5_9,

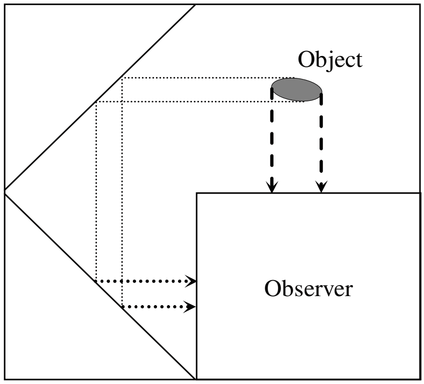


Fig. 1 Reichenbach's cubical world in which an object in the external world casts two shadows observed by the observer

manipulation or intervention. On the other hand, there are well known circumstances in which causal modeling fails. In Section on A Failed Consilience in the Double-Slit Experiment, I argue that the application of Occam's razor to probabilities leads to false predictions in the famous double-slit experiment. The question then arises: Does Occam's razor apply to quantum mechanical probabilities? Or is quantum mechanics (QM) parsimonious with respect to other kinds of entities? This is a question that I attempt to answer affirmatively in sections on Spin Measurements on a Single Electron, The Bell Theorem as Proving a Failed Consilience, and the Consilience of Quantum Mechanics, at least in the context of a concrete example.

The concrete example is interesting because it is similar to the example used by Bell (1964, 1971) to challenge local hidden variable interpretations of QM. I provide a more recent version of Bell's derivation of a false prediction in the section on The Bell Theorem as Proving a Failed Consilience, while in the section on The Consilience of Quantum Mechanics, I show that it is actually very close to the way that QM makes the correct prediction (just replace variables with operators).

Bell's theorem can be viewed as an argument that at least some quantum mechanical phenomena have no causal explanation (van Fraassen 1982), which agrees with the two theses argued in Sections on Consilience in Probabilistic Examples and A Failed Consilience in the Double-Slit Experiment: (1) The constancy, or invariance, of probabilities is an essential component of causal modeling, and (2) the invariance of probabilities leads to false predictions in the double-slit experiment. The most obvious conclusion is that there is no causal explanation of the double slit phenomena, and Bell's argument is important because it is harder to wriggle out of the conclusion. But does QM therefore fail to conform to Occam's razor? Does it fail to describe the world behind the shadows in a unified way? The main goal of this essay is to argue, in at least one concrete example, that it is an exemplary kind cubical world inference, which does conform to Occam's razor.

I have not provided a formal account cubical world inference because there is no such thing in my view. Causal modeling is one kind of cubical world inference, and QM modeling is another, and two are formally quite different. But this is not to say that there is no overarching epistemological theory that covers both cases. In fact, a good informal description of what might be viewed as cubical world inference was published by William Whewell in 1858!

A Whewellian Criterion of Reality

Long before the invention of quantum mechanics, William Whewell (1858) claimed that scientific induction proceeds in three steps: (1) the Selection of the Idea, (2) the Construction of the Conception, and (3) the Determination of the Magnitudes (see Butts 1989, pp. 223–237). In curve-fitting, for example, the selection of the idea is the selection of the independent variable (x) from which we hope to predict the dependent variable (y). The construction of the conception is determined by the choice of the formula connecting the two variables (family of curves). And the determination of magnitudes is what statisticians now refer to as the estimation of the adjustable parameters. Whewell then makes an insightful claim about curve-fitting. If we have a data point (x_0, y_0) and a curve in the x-y plane representing a hypothesis, then the y-value on the curve corresponding to $x = x_0$ is hypothesized to be due to the "signal", while the difference between the y-value on the curve and the observed y-value is the error, or the "noise". But we cannot tell which part is due to the signal and which part is noise by looking at that datum. Rather, we fit a curve through many data points, for then the decomposition into signal and noise that applies to each data point is informed by all the data.

> The philosopher's business is to compare his hypotheses with facts [data], as we have often said. But if we make the comparison with separate special facts [data points], we are liable to be perplexed or misled, to an unknown amount, by the errours of observation; which may cause the hypothetical and the observed result to agree, or to disagree, when otherwise they would not do so. If we thus take the *whole mass of the facts*, and remove the errours of actual observation, by making the curve which expresses the supposed observation regular and smooth, we have the separate facts corrected by their general tendency. We are put in possession, as we have said, of something more true than any fact by itself is. (Whewell, quoted from Butts 1989, p. 227.)

The idea is that the signal is more real than the noise in the sense that it is more permanent and predictable. In fact, when we use the fitted curve to predict new data, we use the curve to make the prediction because the noise is, by definition, random and unpredictable. It's not that noise is unreal. But when it comes to knowledge of the world, it is knowledge of what can be generalized to new cases that is prized. And hypotheses have the ability to generalize to new cases only to the extent that they impose the right conception on the data.

> The particular facts are not merely brought together, but there is a New Element added to the combination by the very act of thought by which they are combined. There is a Conception of mind introduced in the general proposition, which did not exist in any of the observed facts.... The pearls are there, but they will not hang together until some one provides the string. (Whewell, quoted from Butts 1989, pp. 140, 141.)

In the case of curve-fitting, the conceptual string is the curve, but Whewell also insists that the same picture applies to every colligation of facts (which is Whewell's name for 'induction').

Each colligation is confirmed by the prediction of new instances, but the hypothesis is still conjectural, at least with regard to the appropriateness of the conceptions it introduces. If each hypothesis is viewed as a "fact" and if these facts are successfully colligated by successive generalizations, to use Whewell's term, then the resulting hypothesis is "more true" than any of the individual hypotheses from which it is derived. To use Whewell's example, Newton's inverse square law is "more true" than Kepler's laws and Galileo's law of terrestrial motion from which it was induced. Thus, the best theories in the history of science strive towards what he calls the *consilience* of inductions (note the plural). The mark of a good theory lies not in the relationship between the theory and its data *in a single narrow application*, but in the way it succeeds in 'tying together' *separate* inductions. A good theory is like a tree that puts out runners that grow into new trees, until there is a huge forest of mutually sustaining plants. The 'tying together' can be achieved in either of two ways: (a) The theory accommodates one set of data, and then predicts data of a different kind. (b) The theory accommodates two kinds of data separately, and then finds that the magnitudes in the separate inductions agree, or that the laws that hold in each case 'jump together'. When the magnitudes agree, then we have what is more commonly referred to as an agreement of independent measurements.[3] The difference lies only in the historical order in which events take place. In either case, a *consilience of inductions* has taken place, and this leads to a theory that provides a more unified and parsimonious description of the world.[4]

Whewell's defense of the objectivity of knowledge is very different from the logical empiricism that superseded it in the twentieth century. On these views, error is minimized by applying error-free canons of inductive reasoning to error-free data. Whewell's idea is that the colligation of facts is inherently concept-laden, because the facts are combined by a new conception of mind. We need to distinguish between signal and noise, tentatively at first. But subsequently, there is a process of objectification in which the inductive hypotheses are colligated at a higher level of generalization. More importantly, the criterion by which we judge the reality of postulated quantities is *empirical* if it involves the agreement of independent measurements of postulated quantities. A careful examination of other examples in the history of science, such as Newton's argument for universal gravitation (Forster 1988; Harper 2002), shows that this empirical element has played a fundamental role in triggering a revolutionary change in the way we view the world. If it has happened before, it can happen again.

The purpose of this essay is to examine the consilience of colligations in quantum mechanics. In this respect, I argue, the new physics is the same as the old. The difference is that the resulting picture of reality does not conform to our usual common sense picture of the world. The reason is that consilience cuts both ways. The right kind of consilience can support a common sense view of the world. But when consilience fails to occur, we must seek out new explanations that do achieve consilience.

Consilience in Probabilistic Examples

Quantum mechanics is essentially a probabilistic theory. So to understand how the notions of colligations and consilience apply to it, we should first look at classical examples of probabilistic modeling.

Consider a very simple example – the jackpot machine. The machine has two input states: Either one euro coin is placed in the machine, or two euro coins are placed in the machine, then a handle is pulled. The output state is either 'win' or 'lose'. For the sake of concreteness, suppose 'win' refers to the event that machine delivers 10€ as a 'jackpot', and a loss leads to a zero payout. In this example, the input states do not *determine* unique output states. The 'mechanism' built into the machine is best described in terms of the constancy or invariance of *probabilities*.

To make this point, introduce two variables, X and Y. Upper case letters are used in statistics to denote variables that have probabilities associated with their possible values. Such variables are called *random variables*. X represents the input state of the machine in some particular trial of the experiment, while Y represents the output state in the same trial. The possible states of the machine can be represented by assigning arbitrary numerical values to these variables. Let $X = 1$ denote the event that only 1€ is placed in the machine before the handle is pulled, while $X = 2$ denotes the event that 2€ are placed in the machine. $Y = 0$ denotes the event that there is no payout, while $Y = 0$ denotes the event that the jackpot (of 10€) is paid out.

The standard 'forward' causal model says:

$$P(\text{win}|1€) = \alpha, \text{ and } P(\text{win}|2€) = \beta, \text{for all trials } i.$$

That is, different trials have something in common; namely, that the values of the forward conditional probabilities are the same in each case. Note that the model postulates constant values for *all* forward probabilities because

$$P(\text{loss}|1€) = 1 - \alpha, \text{ and } P(\text{loss}|2€) = 1 - \beta, \text{for all trials } i.$$

These four probabilities are *postulated* by the model. They count as *theoretically* 'entities'. Their measurement, or estimation, is determined from the available data in the same way that any theoretical quantity is measured. Consider a typical set of data. Suppose that Alice plays the machine 200 times, and we record the input and output state on each trial. The data consist in a sequence of data 'points', which come in four kinds: (1,0), (1,1), (2,0), or (2,1). Because the model says that the temporal order of the trials does not matter, the data are adequately recorded in the table of the observed frequencies.

(Alice)	Loss	Win
1€	90	10
2€	80	20

Notice that Alice plays the machine with 1€ half the time and with 2€ half the time (100 trials each). Out of all the times she plays the 1€ version of the game, she wins the jackpot ten times, thus earning 100€, which is the same amount that she paid to play. Out of all the times she pays 2€ to play, she wins twice as often, and earns a total of 200€ (twice as much). But she paid twice as much to play, so she still earned the same as what she paid to play. On the basis of the data, the machine appears to be fair.

Whewell's three steps in the colligation of facts apply to this example in the following way. Step 1 consists in the selection of X and Y as the relevant quantities to be considered. Step 2 introduces the formula, which in this case is probabilistic in nature. It introduces a family of probability distributions parameterized by the adjustable parameters α and β. These determine the Conception. In this example, the conception is probabilistic in nature. Step 3 is the determination of the magnitudes α and β from the data. In contemporary statistical theory, this is achieved by the method of maximum likelihood estimation (MLE).

To understand how MLE works, first note that each pair of values assigned to α and β picks out a particular probabilistic hypothesis in the model. The fit of each hypothesis with the data is defined by its likelihood, which, by definition, is the probability of the data given the hypothesis (this should not be confused with the probability of the hypothesis given the data, which is a distinctly Bayesian concept). The greater the likelihood of a hypothesis (the more probable it makes the data) the better the hypothesis fits the data. The hypothesis that fits best is, by definition, the maximum likelihood hypothesis. For arbitrary values of α and β, the likelihood is:

$$\text{Likelihood}\ (\alpha, \beta) = (1-\alpha)^{90}\alpha^{10}(1-\beta)^{80}\beta^{20}.$$

The probabilities are multiplied together because each trial is probabilistically independent of all the others (according to the model) in the same way that coin tosses are independent. Mathematically speaking, maximizing the likelihood is the same as maximizing the log-likelihood.

$$\text{log-Likelihood}\ (\alpha, \beta) = [90\log(1-\alpha) + 10\log(\alpha)] + [80\log(1-\beta) + 20\log(\beta)].$$

Again, the terms in the square brackets can be maximized separately by differentiating with respect to α and β and putting the resulting expressions equal to zero. (You need to know that $d\log x/dx = 1/x$.) After multiplying by the factors $\alpha(1-\alpha)$ and $\beta(1-\beta)$, respectively, the equations simplify to:

$$-90\alpha + 10(1-\alpha) = 0 \text{ and } -80\beta + 20(1-\beta) = 0.$$

These equations yield the estimates: $\hat{\alpha} = 0.1$ and $\hat{\beta} = 0.2$. Accordingly, the theoretically postulated probabilities are estimated by the natural relative frequencies in the data, just as one would naïvely expect.

The question is whether the 'forward' model, which is the standard model, has evidence in its favor that the 'backward' model does not. The backward model seeks to predict values of X from values of Y (in the same probabilistic sense of 'prediction'). In other words, it postulates backward probabilities as follows:

$$P(2€|\text{loss}) = \gamma, \text{and} P(2€|\text{win}) = \delta, \text{for all trials } i.$$

Using the same methods as before, the parameters of the backward model are estimated by the corresponding relative frequencies in the data. The estimated values are: $\hat{\gamma} = 0.47$ and $\hat{\delta} = 0.67$. There is nothing that points to any empirical difference between the models. In fact, they cannot be compared with regard to their fit to the data because they tackle very different prediction tasks. The fit of the forward model is measured in terms of Y-values, while the fit of the backward model is measured in terms of X-values. In a sense, the two models are incommensurable.

So, why don't the two models happily co-exist? There is a sense in which they do co-exist, for the forward model is capable of making backward predictions if it is provided with information about the relative frequency of 1€ versus 2€ trials. With this information, together with the estimated forward probabilities, the forward model can calculate the backward probabilities, and gets the same answer as backward model. This is because both models accommodate the same table of data. But this only serves to deepen the puzzle. If both models can be seen as predictively equivalent, why interpret one as causal and not the other?

This is the familiar puzzle about cause and correlation: The evidence for causation cannot be exhausted by the correlations in the data, for correlations are symmetric, while causation is not. Either there is no additional empirical evidence, in which case causal inference is based on non-empirical criteria (or psychological habit!), or else there is other evidence that breaks the symmetry.

One solution is to look at how the models predict data of a different kind. Suppose that Bob plays the jackpot machine, and he happens to play with 2€ twice as often as with 1€. Given that the forward model is true, the frequency that Bob wins will conform to the same forward probabilities, modulo a fluctuation in the data due to sampling errors. The sampling fluctuations are not relevant to this discussion, so imagine that there are no sampling errors. Then data from Bob's trials are described by the natural frequencies in the table.

(Bob)	Loss	Win
1€	90	10
2€	160	40

Given the way that the example was set up, it is hardly surprising that the independent measurements of α and β agree. The key question is whether the same is true for the backward model. If it is not, then the symmetry between the models is broken.

A simple calculation shows that the Bob's estimates of the parameters of the backward model are $\hat{\gamma} = 0.64$ and $\hat{\delta} = 0.80$, which are quite different from the previous values, which were $\hat{\gamma} = 0.47$ and $\hat{\delta} = 0.67$. Therefore, the backward model fails the test of consilience, and the symmetry between the forward and backward model is broken on empirical grounds (Forster 1984; Sober 1994; Arntzenius 1997).

The symmetry between the models is restored if Alice's and Bob's data are pooled – in fact, the models are always symmetric with respect to any single data set, as was already shown by considering Alice's data. The evidence is relational; and the consilience of inductions is a relation between different colligations of facts.

A Failed Consilience in the Double-Slit Experiment

The postulation of hidden variables is a way of interpreting the probabilities of quantum mechanics as 'measures of ignorance'. Einstein, for example, believed that a future physics would reveal the existence of such hidden variables, and the decay of radioactive particles, for example, could be predicted exactly. Bohr, on the hand, thought that quantum physics was complete in the sense that quantum probabilities are here to stay.

At the time of the debate, a common example was the double-slit experiment. Consider a particle of light (photon) that leaves the light source and travels through either slit A or slit B (and not both). We may represent this event in terms of a random variable X, which can have one of two values, x_A and x_B.[5] For our purposes, it doesn't matter what numerical values we use. Now introduce a second random variable Y such that $Y = 1$ if the particle is detected in some region C, and 0 otherwise. In a more natural shorthand notation, let A stand for the event $X = 1$, B of for the event $X = -1$, and C is the event $Y = 1$. The probability of C given A is written $P(C|A)$. Similarly, $P(C|B)$ is the probability that it arrives at C given that it is passes through slit B. $P(A)$ is the probability that the particle passes through slit A given that it passes through any slit. Then the probability of C (given that it arrives somewhere on the screen) is, according the axioms of probability:

$$P(C) = P(A)P(C|A) + P(B)P(C|B).$$

	C	Not-C
A	40	10
Not-A	20	30

It is worthwhile working through the proof of this theorem, because it introduces some elementary concepts of probability theory that are controversial in QM. Let's suppose that the particle passes through slit A or B, but not both. Then B is logically equivalent to the event not-A. Suppose that 50 particles pass through slit A and 50 particles pass through slit B. Of the 50 particles that pass through slit A, 40 arrive at C. Of the 50 particles that pass through slit B, 20 arrive at C. Therefore, a total of 60 out of 100 particles arrive at C. In symbols, $P(A) = 50/100$, $P(C|A) = 40/50$, $P(B) = 50/100$. $P(C|B) = 20/50$. Therefore,

$$P(A)P(C|A) + P(B)P(C|B) = \frac{50}{100}\frac{40}{50} + \frac{50}{100}\frac{20}{50} = \frac{60}{100} = P(C)$$

$P(A\&C)$ is the probability that the particle passes through slit A *and* arrives at C. According to the table, this joint probability is $40/100$. The argument assumes that joint probabilities, such as $P(A\&C)$, exist.

So far, there is no problem. But now postulate that the probabilities $P(C|A)$ and $P(C|B)$ do not depend on whether the other slit is open or closed. This is often called a *locality assumption* because it assumes that what happens to a particle passing through one slit is unaffected by what is happening non-locally (at the other slit in this case). It is also an invariance assumption – it is an attempt to unify the phenomena of the single-slit and double-slit experiments. It is might be viewed as an application of the principle that "Probabilities are not to be multiplied beyond necessity."

Unification is important because it increases predictive content. But in this case it leads to the *false* prediction that the double-slit pattern is the average of the two single slit patterns. If the slits are a certain distance apart, then the double-slit pattern has its brightest spot at C (see Fig. 2), whereas any average of the single slit patterns has the brightest spots directly in front of the two slits. The invariance of the probabilities provides a *potential* consilience of inductions, but when the consilience fails, the assumptions on which it is based are called into question.

There are ways of resisting this conclusion. The first response is that different photons are interacting with each other after they pass through the slits. This possibility is highly implausible in light of the fact that the interference pattern is exactly the same if the intensity of the light is so low that only one photon passes through the slit at a time.

Another possibility is that there is some kind of continuously emitted pilot wave that guides the particles to their appropriate destinations. The pilot wave is affected by whether the second slit is open or closed. This hypothesis is ad hoc if there is no way of independently detecting the existence of the pilot wave. But it doesn't lead to any false predictions. To that extent it solves the problem. But, ultimately, it is unsuccessful if it does not lead to any new consilience.

The argument that *probabilities* should be invariant across the single-slit and double-slit experiments is based on the assumption that the particles pass through one slit or the other in every instance. It is an important prediction of quantum mechanics that if a precise trajectory is experimentally determined, then the classical

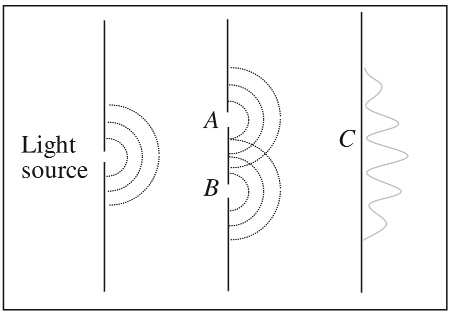


Fig. 2 In the double-slit experiment, the assumption that individual photon travel through either slit A or slit B leads to the false prediction that double-slit pattern is the sum of two single-slit patterns

prediction *is* correct – that is, in this case the pattern on the screen will be the sum of the two single-slit patterns. For example, electrons can be detected going through a particular slit by the electric current they induce when they pass through a wire loop. So if we place wire loops behind the two slits, then the sum of the two single slit patterns will be observed.

The problem with the causal explanation is that it *sometimes* makes false predictions. So, how does the quantum mechanical formalism succeed where the causal account fails? The causal model unified the phenomena by assuming that the *probability distributions* in the two single slit experiments *add* together to produce the probability distribution in double slit experiment. Quantum mechanics replaces the additivity of probabilities with the additivity of wave functions. Let's introduce a random variable Y, where y denotes an arbitrary value of Y, to represent the possible points on the screen at which the particle may be detected. Now consider the single slit experiment in which slit B is closed. Suppose a quantum mechanical model entails a wave function that does not depend on time, and has a complex number $\psi_A(y)$ associated with each point on the screen. Then the model implies that the probability that a particle is detected near the point y is proportional to $|\psi_A(y)|^2$. Note that the square of the magnitude of complex number is a non-negative real number. Similarly, the probability of a particle landing near y is $|\psi_B(y)|^2$ in the other single slit experiment. Then the probability of a particle landing near the same point in the double slit experiment is:

$$\left|\frac{1}{\sqrt{2}}\psi_A(y) + \frac{1}{\sqrt{2}}\psi_B(y)\right|^2 = \frac{1}{2}|\psi_A(y)|^2 + \frac{1}{2}|\psi_B(y)|^2 + \text{interference terms.}$$

If the interference terms are zero, then the prediction is that same as the hidden variable prediction. QM allows for the additivity of probabilities, but the additivity, or superposition, of wave functions is more fundamental. The QM model succeeds in unifying a wide range of disparate phenomena. It leads to a consilience within a vast network of inductions.

In summary: QM modeling is very different from causal modeling because it uses operators in place of variables, and uses wave functions (state vectors) rather than probabilities to represent the constancies of nature. Nevertheless the conceptual novelties of QM produce an impressive consilience of a variety of inductions, which is the strongest kind of evidence that any scientific theory can have.

Spin Measurements on a Single Electron

In the quantum mechanical model of electron spin, there is a QM spin observable corresponding to spin in every direction in three-dimensional space. We should not try to read too much into the word 'spin' in quantum mechanics. For us, it is just a name of a new kind of QM property.

There are really only two facts that you need to understand about quantum mechanical spin. The first is that if an electron passes through a non-linear magnetic

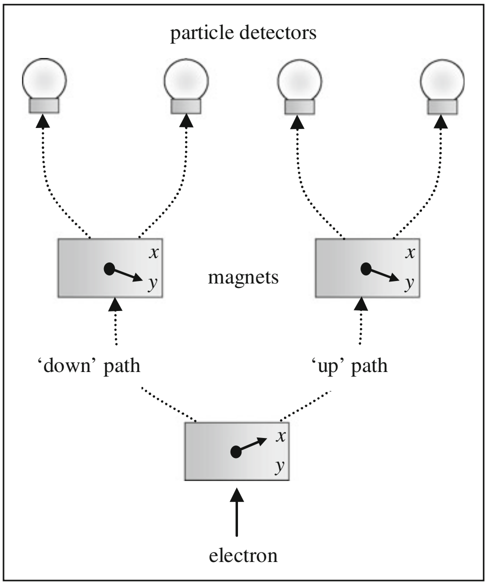


Fig. 3 Sequential spin measurements on a single election that lead to four possible outcomes

field produced by a Stern–Gerlach magnet, then the exiting electron will be detected in one of two possible paths, the 'up' path or the 'down' path, provided that detectors are placed there. We shall always assume that the electron is always traveling in the z direction and all Stern–Gerlach magnets are oriented in a direction perpendicular to z (for example, x or y).

In a *sequential measurement* a single election is first 'prepared' by passing it through one magnetic field, after which it is directed through a second Stern–Gerlach device. We need three devices in all, and there are four possible exit paths, and the particle is detected in exactly one path by the detectors placed there (see Fig. 3). When a particle is detected, it is destroyed, so that it cannot be subjected to further measurement.

If the particle is detected by the Geiger counter placed in the 'down path of the magnet placed in the 'up' path of the first magnet, then we say that the outcome of the experiment is 'up–down', or $+-$. In a classical framework, we infer from this outcome that if detectors had been placed in the exits paths of the first magnet, it would have been detected in the 'up' path. We may represent this conclusion by assigning a number to a variable. Counterfactual assumptions of this kind are controversial in QM, but it does provide a way of assigning numbers to variables on the basis of observed experimental outcomes. The legitimacy of the *interpretation* remains an open question.

Assume that the first magnet is oriented in the x-direction, and introduce $X_0 = 1$ to mean that if the detectors had been placed directly in the exit paths, without any intervening device, then it would have been detected in the 'up' path. Now consider the fact that the outcome was 'up–down' rather than 'up–up'. By convention, the statement $Y_1 = -1$ means that the electron would be detected in the 'down' path after passing through a device oriented in the y-direction. Further suppose that the second device was actually oriented in the y-direction. Then we infer that $Y_1 = -1$. Therefore, from the experimental outcome, we infer that $X_0 = 1$ and $Y_1 = -1$. There are four possible conclusions drawn from four possible outcomes.

From repeated trials of this experiment we discover that each of the four outcomes occurs equally often. This statistical fact is represented in terms of the probabilistic statements

$$P(X_0 = 1) = \tfrac{1}{2} \text{ and } P(Y_1 = 1|X_0 = 1) = \tfrac{1}{2} = P(Y_1 = 1|X_0 = -1),$$

which state that the two variables are probabilistically uncorrelated. Various other facts emerge as well, such as the fact that the statistics are insensitive to the distance separating the second magnet from the first or how far the detectors are placed from the second magnet. By varying the orientations of the magnets, we also discover that the correlation between the variables depends only on the angle between them. When the magnets are oriented in the same direction we say that the electron is subjected to a repeated measurement. In that case, the variables are perfectly correlated. That is, $X_1 = 1$ if and only if $X_0 = 1$, or in probabilistic terms,

$$P(X_1 = 1|X_0 = 1) = 1 \text{ and } P(X_1 = 1|X_0 = -1) = 0.$$

This hidden variable representation of the observational facts can be extended to an arbitrary number of sequential measurements. For example, if we 'measure' the values of X_0, Y_1 and X_2, then we discover that the statistical relationship between X_0 and Y_1 is the same as before. It is not changed by the third measurement. Moreover, the statistical relationship between the last two variables is independent of the inferred value of the first variable. In symbols,

$$P(X_2 = 1|Y_1 = 1, X_0 = 1) = P(X_2 = 1|Y_1 = 1, X_0 = -1) = \tfrac{1}{2},$$

and so forth. This is a kind of Markov condition, according to which proximal causes "screen off" distal causes. But the relationship between the first and the third variables is changed by the second measurement. In particular, we find that $P(X_2 = 1|X_0 = 1) = \frac{1}{2}$ and $P(X_2 = 1|X_0 = -1) = \frac{1}{2}$, which is different from the relationship between X_0 and X_1. X_0 and X_2 are uncorrelated, whereas X_0 and X_1 were correlated. In fact, the relationship between X_0 and X_2 depends on the existence of an intermediate device and its orientation. For example, if the intermediate measurement were in the x-direction, then the values of X_0 and X_2 would be the same (perfect positive correlation). All this implies that the probability distributions depend on features of the *total measurement setup*. Once this caution is in place, there are no false predictions that are automatically derived from the hidden variable representation of sequential spin measurements. The false predictions rely on assumptions about how the probabilities in different experimental contexts relate to each other.

The experimental facts concerning sequential spin ½ measurements might be summarized in the following way. Suppose that a measurement in the u-direction is followed immediately by a measurement in the v-direction, with associated variables U and V. Let θ be the angle between u and v. Then

$$P(V = 1|U = 1) = \cos^2(\theta/2) \text{ and } P(V = 1|U = -1) = \sin^2(\theta/2).$$

Notice that these forward probabilities are invariant, as required of the very best causal models. In fact, this quantum phenomenon would be ideal for building a jackpot machine because the probabilities are precisely controlled by the orientations of the magnets and nothing else. However, is the underlying mechanism really causal? Any model must be subjected to ever broadening standards of consilience. If some new-fangled QM model preserves the consilience of sequential spin phenomena and extends its predictions to a broader domain, then it trumps the causal model.

This is exactly what happens. Hidden variable models that make further assumptions about the irrelevance of what happens non-locally in space and time make predictions that contradict the predictions of QM, as Bell (1964) first proved. Moreover, there is now a fairly broad consensus that the predictions of local hidden variable theories are false. Therefore, *local* hidden variable models are false. Since a local causal model entails the probabilistic assertions of the local hidden variable model, those causal models are also false. Since the locality assumption is an attempted consilience of inductions, this another example of a failed consilience. At least, this is the way I will present the example.

The Bell Theorem as Proving a Failed Consilience

Since Bell's argument was originally presented as a proof against the soundness of the Einstein Podolsky and Rosen (1935) argument for the incompleteness of QM, I will summarize that argument. In the original EPR thought experiment, two particles are prepared in a QM state, and they fly apart to opposite ends of the universe. The EPR argument is summarized as follows.

1. EPR criterion of reality: If, without in any way disturbing a system, we can predict with certainty (i.e. with probability equal to unity) the value of a quantity, then there is an element of reality corresponding to that quantity.
2. EPR example: There is a QM system consisting of particles I and II, where that P and Q are incompatible observables pertaining to system II, such that P can be predicted with certainty if quantity A on particle I is measured and that Q can be predicted with certainty if quantity B on particle I is measured.
3. Locality: The measurements of A or B on particle I do not disturb the state of particle II or affect the choice of measurement on particle II in any way.
4. Particle II has property P (or $\bar{P}$), since we *could* measure A and predict P (or $\bar{P}$), and infer that P (or $\bar{P}$) corresponds to an element of reality. Therefore,

Particle II has property P (or $\bar{P}$) *whether or not A is actually measured*. Similarly, Particle II has property Q (or $\bar{Q}$) whether or not B is measured. This conclusion cannot be reached without the locality assumption.
5. Therefore there exists elements of reality corresponding to P and to Q, at the same time.
6. P and Q are incompatible properties in QM, which means that QM never assigns them precise values (i.e., values predicted with probability 1) at the same.
7. Therefore QM is incomplete. That is, there exist elements of reality not represented in QM. These elements of reality are represented by hidden variables, which have to be added to the usual formalism of QM. Finally, it follows from the locality assumption that these hidden variables are *local*.

The shocking fact is that the EPR argument is *provably* unsound, as was originally shown by Bell in 1964. This section presents a perspicuous version of the proof (Greenberger et al. 1989; Mermin 1990), which helps provide some insights into the physics of the example and will provide an excellent illustration of how the QM formalism succeeds where hidden variables fail (the topic of the next section).

Bell's argument takes the form of a reductio ad absurdum proof: It assumes that the local hidden variable theory is true, then allows it to accommodate one set of experimental facts, and then shows that it makes a false prediction in another set of experimental facts. Therefore, the hidden variable theory is false.

Here is a dry run. Suppose that couples are interviewed in a psychology experiment. Each partner is taken from the waiting room to two sound proof interview rooms, and each is asked one of two questions according to separate coin toss. Nobody knows in advance which question will be asked, although everyone knows that it will be question X or question Y. Suppose that the value of X determines the answer to the question "Are you a Democrat?" and Y determines the answer to the question "Are the Greenbay Packers your favorite football team?". The $+1$ means YES and -1 means NO. Suppose that the experimental facts show that when both partners are asked the same questions, they answer either yes–yes or no–no.

That leaves open whether they would give the same answers if they were asked different questions. But it is not completely open. It is not possible to get a perfect agreement when partner 1 is asked X and partner 2 is asked Y, but a perfect disagreement when the questions are switched. To prove this, let's introduce X_1 to be the variable whose value determines the answer given when partner 1 is asked question X. Similar meanings are assigned to Y_1, X_2 and Y_2. Recall, to get the first set of facts right, the instructions must satisfy the constraints $X_1 = X_2$ and $Y_1 = Y_2$. That is, the same answers are given when the same questions are asked. What this implies is that we only need to specify the values of X_1 and Y_1 in order to determine the full instruction set for both partners. But now suppose that when X is asked of partner 1 and Y is asked of partner 2, then the answers are also the same. Then $X_1 = Y_2$. But since $Y_1 = Y_2$, this implies that $X_1 = Y_1$. Now there are only two possible instructions sets: Either all answers to all questions are YES or all the answers to all questions are NO. This *predicts* that when Y is asked of partner 1 and X is asked of partner 2, then the answers must also be the same.

But suppose that experimenters find that this prediction is false! It is possible that the couples are playing a prank? To succeed, partner 1 is instructed to answer YES to both questions, and partner 2 knows this. Now, suppose partner 2 is asked X. The problem is that answer she gives must depend on the question that her partner is asked. If her partner is asked X she must answer YES. If her partner is asked Y, she must answer NO. She can't get this right (all the time) unless she knows what question her partner has been asked. But she doesn't know what question her partner is being asked. That is the locality assumption. The standard causal story stands refuted.

Here is the QM version of the same argument. Three electrons fly apart towards three widely separated Stern–Gerlach magnets, labeled 1, 2, and 3 (Fig. 4). Each magnet is aligned in one of two directions (x or y) orthogonal to the path of the incoming electron, and each device contains two particle detectors, one placed in the 'up' path and one in the 'down' path exiting the magnet, such that if the electron is detected in the 'up' path, the light bulb attached to the device flashes red and if it is detected in the 'down' path, the same light bulb flashes green. So, each bulb flashes either red or green. We may say that the outcome is spin in the direction at which the magnet is oriented is 'up' if the light is red and 'spin-down' if the light is green.

Consider the first set of experimental facts: When any two of the measurement devices are set to y and the third is set to x (that is, for settings y–y–x, y–x–y, or x–y–y) there is always an odd number of red lights flashing in every trial of the experiment – that is, either all three lights flash red or one light flashes red and the other two flash green.

A local hidden variable model can accommodate these experimental facts in the following way: Suppose that each particle, after separation from the others, carries with it a set of properties that determine which light bulb will flash for every possible

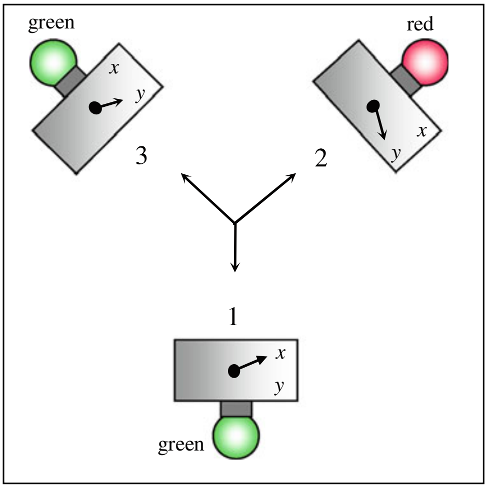


Fig. 4 The GHZ version of Bell's experiment

settings of the device it enters. Let's represent the property that particle 1 would cause the red bulb to flash if the device 1 were set to position x by $X_1 = +1$, and the property that the green bulb would flash were device 1 set to position x by $X_1 = -1$. According to the hidden variable story, the particle has the property $X_1 = +1$ or the property $X_1 = -1$, but not both.

Note that while the value of the variable *determines* the experimental outcome, the variable is not being used to represent the observable outcome. The existence of these variables is *postulated by the theory* to 'explain' the observed outcome. The theory assumes that the hidden variables have values even when they are not measured.

Similarly, let $Y_1 = +1$ and $Y_1 = -1$ represent the two properties that determine the outcome when the measuring device 1 is set at y. Then, a particle heading towards device 1 will have exactly one of four possible sets of properties, which Mermin (1990) refers to as "instruction sets": Either $\{X_1 = +1,\ Y_1 = +1\}$, $\{X_1 = +1,\ Y_1 = -1\}$, $\{X_1 = -1,\ Y_1 = +1\}$, or $\{X_1 = -1,\ Y_1 = -1\}$. Similar hidden variable states are assigned to particles 2 and 3. There are therefore six hidden variables that collectively play the role of a common cause, and these may adjusted in whatever way is needed to accommodate the facts.

Note that in a single run of the experiment, a measurement device cannot be oriented in two directions simultaneously, so we cannot determine all the hidden variable values by direct measurement. For instance, if we see the bulb flash red when device 1 is set to y, then we would only know that the instruction set was either $\{X_1 = +1,\ Y_1 = +1\}$ or $\{X_1 = -1,\ Y_1 = +1\}$. Some of the variables are always hidden. But their existence still has empirical consequences. What needs to be accommodated is that fact that the outcome for the third particle must be R if and only if first two outcomes are either $R-R$ or G–G. The constraints that are necessary and sufficient to accommodate all three regularities are:

$$X_1 Y_2 Y_3 = 1,\ Y_1 X_2 Y_3 = 1,\ \text{and}\ Y_1 Y_2 X_3 = 1. \tag{1}$$

For example, if the setting is y–x–y, then the second equation tells us that if the outcome for particles 1 and 3 is R–R, then $Y_1 = 1 = X_2$, and therefore $Y_3 = 1$, and so the outcome for particle 3 will be R.

If the theoretical laws in (1) are assumed to apply in all experimental situations, then the following deductive consequence of the laws holds in all situations. Multiply the three equations in (1) together, and simplify using $Y_1^2 = Y_2^2 = Y_3^2 = 1$, to obtain

$$1 = (X_1 Y_2 Y_3)(Y_1 X_2 Y_3)(Y_1 Y_2 X_3) = X_1 X_2 X_3.$$

The equation, $X_1 X_2 X_3 = 1$, implies that there will also be an odd number of red flashes when all three devices are set to x-x-x. This prediction is dramatically different from the prediction made by QM, which predicts that there must be an *even* number of red flashes in this experiment! QM is right and the hidden variable prediction is wrong![6]

The Consilience of Quantum Mechanics

QM models not only make the correct predictions about entangled states, but they do so *on the basis of what is known* about sequential measurements. As Whewell might have put it, the QM induction of sequential data explains and predicts facts of a different kind, and it does so by superimposing a new conception on the facts. The remainder of section is designed to deepen our understanding of the colligation facts in QM.

Consider the 'empirical fact' $P(X_1 = -1|X_0 = 1) = 0$. QM introduces vectors x^+ and x^- to represent the states $X_0 = 1$ and $X_1 = -1$, respectively, and then derives the probability $P(x^-|x^+) = 0$ from the mathematical properties of vectors. The probability is 0 because the projection of the vector x^+ onto the vector x^- is 0, which is to say that x^+ and x^- are orthogonal vectors. The same argument applies to the two states associated with an arbitrary direction of measurement v^+ and v^-, where the physical angle between the two directions is θ. The question is whether there exists a QM model in which all four vectors are in the same complex 2-dimensional quantum mechanical state space? If there exists such a unified model, then the four vectors would appear as drawn in Fig. 5, for some angle ϕ. In that case, any vector in the space can be written as a linear combination of x^+ and x^-. That is, all states would be superpositions of the states x^+ and x^-. Superposition is therefore a unifying conception in QM.

The crucial question is whether it is possible to adjust ϕ so that the 'cross' probabilities, $P(v^+|x^+)$ and $P(v^+|x^-)$, are related in the right way to the physical angle θ. To answer this question, we need to understand more about how probabilities are calculated in QM.

Clearly, we want to calculate probabilities in such a way that $P(v^+|x^+) + P(v^-|x^+) = 1$, since any electron prepared in the x^+ state will go 'up' or 'down' after exiting a Stern–Gerlach magnet oriented in any direction. A trivial instance of this is $P(x^+|x^+) + P(x^-|x^+) = 1$. The vector space formalism attributes this probability to the geometric fact that projecting x^+ onto x^+ leaves x^+ unchanged. But how do we extract the number 1 from this geometric fact? QM says in general that $P(v^+|x^+)$ is the dot product of the vector x^+ and the projection of x^+ onto v^+. Don't ask why this works – nobody really knows, or at least, nobody agrees on the answer. In the special case in which $v^+ = x^+$, this is the dot product of x^+ times

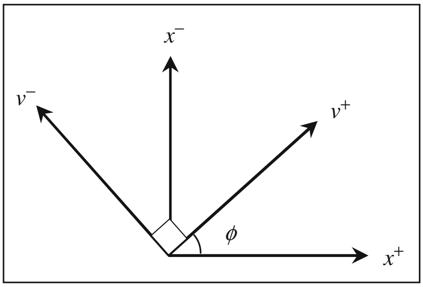


Fig. 5 In QM, the transition probabilities between states can be generated from the geometrical relationship between vectors

x^+, which is just the square magnitude of the vector x^+. Clearly, this 1 if and only if x^+ is a unit vector, so all state vectors in QM are unit vectors.

Now reconsider $P(v^+|x^+)$. Since all vectors are unit vectors, the dot product of x^+ and the projection of x^+ onto v^+ is equal to $(\cos\phi)\, v^+$. The dot product of this with x^+ is

$$\left(\cos\phi\, v^+\right) \cdot x^+ = \cos\phi\,(v^+ \cdot x^+) = \cos^2\phi.$$

Clearly, we will match the empirically known probabilities if and only if $\phi = \theta/2$. Incidentally, this is why it's called spin 1/2 in QM.

In place of arithmetic variables, QM makes use of geometric objects. What's remarkable is that these geometric objects have properties that not only colligate one kind of phenomena, but extend also to disparate classes of facts.

Vectors may be represented as column matrices or as row matrices; we need a notation that reflects the difference. Instead of x^+, we shall write $\left|x^+\right\rangle$ for the column vector, while $\left\langle x^+\right|$ stands for the corresponding row vector. The dot product of row vectors $[u_1 \quad u_2]$ and $[v_1 \quad v_2]$ is equal to $u_1 v_1 + u_2 v_2$, which is the result of multiplying a row vector with a column vector. That is why the dot product of u^+ and v^+ is written $\left\langle u^+ \mid v^+\right\rangle$. The dot product of a vector with itself is just the squared magnitude of the vector; in symbols, $\left\langle x^+ \mid x^+\right\rangle = \left|\left|x^+\right\rangle\right|^2$. In Dirac's notation, $\left\langle x^+ \mid x^+\right\rangle$ is a bracket (bra-ket) and hence $\left\langle x^+\right|$ is called a bra-vector and $\left|x^+\right\rangle$ a ket-vector.

Since the vector space is 2-dimensional, $\left|x^+\right\rangle$ and $|x^-\rangle$ form a complete set of mutually orthogonal unit vectors, called an orthonormal basis. That is, every column vector can be expressed as a linear combination (superposition) of the vectors $\left|x^+\right\rangle$ and $|x^-\rangle$. Similarly, the bra- vectors $\left\langle x^+\right|$ and $\langle x^-|$ form an orthonormal basis of the dual space of row vectors. In this basis,

$$\left|x^+\right\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } |x^-\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

and an arbitrary vector can be expressed as the superposition $|v\rangle = \cos\phi\left|x^+\right\rangle + \sin\phi\,|x^-\rangle$, which is a unit vector because $\cos^2\phi + \sin^2\phi = 1$. But what is the operator, denoted by P_v that projects other vectors onto $|v\rangle$? Any projection operator must have the following properties:

$$\mathrm{P}_v\,|v\rangle = |v\rangle\,,\ \langle v|\,\mathrm{P}_v = \langle v|\,,\ \text{and}\, \mathrm{P}_v\left|v^\perp\right\rangle = |0\rangle\,,\ \left\langle v^\perp\right| P_v = \langle 0|\,,$$

if $\left|v^\perp\right\rangle$ is orthogonal to $|v\rangle$. Also, it is easy to see that

$$P_v = |v\rangle\,\langle v|$$

has the required properties, because $|v\rangle\,\langle v \mid v\rangle = |v\rangle$, and so on. So, if $v = [\cos\phi \quad \sin\phi]$, then associated the matrix that projects any vector onto v is

$$P_v = \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix} \begin{bmatrix} \cos\phi \ \sin\phi \end{bmatrix} = \begin{bmatrix} \cos^2\phi & \cos\phi\sin\phi \\ \cos\phi\sin\phi & \sin^2\phi \end{bmatrix}.$$

In particular,

$$\mathrm{P}_{x+} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \mathrm{P}_{x-} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \text{ and } \mathrm{P}_{y+} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \mathrm{P}_{y-} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

where we have used the fact that $\phi = \theta/2 = \pi/4$ and $\cos\pi/4 = \sin\pi/4 = 1/\sqrt{2}$.

All QM observables are definable in terms of the projection operators. To find the observable corresponding to 'spin' in the x-direction, we construct an operator that has the right mean value in every state. If spin 'up' is associated with the number $+1$ and spin 'down' associated with -1, then the expected value of the spin observable σ_x in state v is

$$(+1)P(x^+|v) + (-1)P(x^-|v) = \langle v|\,\mathrm{P}_{x+}\,|v\rangle - \langle v|\,\mathrm{P}_{x-}\,|v\rangle = \langle v|\,\mathrm{P}_{x+} - \mathrm{P}_{x-}\,|v\rangle\,.$$

This proves that the operator that has the correct mean value is

$$\sigma_x = P_{x+} - P_{x-} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

This is one of the three Pauli spin matrices (modulo a constant factor). The second Pauli matrix is

$$\sigma_y = P_{y+} - P_{y-} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

It is now trivial to prove that anti-commutation of σ_x and σ_y;

$$\sigma_x\sigma_y = -\sigma_y\sigma_x.$$

This is property of spin observables that leads to the correct predictions in the three-particle experiment described in the next section.

The important lesson of this section is that anti-commutation relation follows necessarily from a simple QM model that introduces just one adjustable parameter, the angle ϕ between vectors, which fits a huge variety of facts when we set $\phi = \theta/2$.

I will now *derive* the correct prediction in the GHZ example from quantum mechanical principles. Instead of using *variables*, quantum mechanics associates spin *observables* to each particle for every possible orientation of the Stern–Gerlach magnet. So, for example, the observable ${\sigma_x}^1$ replaces the hidden variable X_1, and the observable ${\sigma_y}^3$ replaces the hidden variable Y_3, and so on.

There is another technical point that will prove useful here. Let us define an observable to be *dispersion-free* in state ψ if and only if the probability of all outcomes is zero except one, which has probability one. In classical physics, a variable is dispersion-free if and only if its variance is zero. Similarly, a QM observable $\hat{A}$ is dispersion-free in state ψ if and only if its variance is zero, where the variance is defined as the expected value of $\left(\hat{A}-a\right)^2$, where a is the mean value of the observable in state ψ.[7] It is relatively easy to prove that $\hat{A}$ is dispersion-free if and only if $\hat{A}\,|\psi\rangle = a\,|\psi\rangle$ for some real number a. In words, $\hat{A}$ is dispersion-free in state ψ if and only if ψ is an eigenstate of $\hat{A}$. A proof is found in Khinchin (1960, pp. 54–55).

For example, if the system were in a $+1$ eigenstate of the observable ${\sigma_x}^1$, then we would predict with probability 1 that device 1 will flash red. If the system is in the -1 eigenstate of the same observable, then the green light will flash, and so on. If the system is not in an eigenstate of ${\sigma_x}^1$, then the probability that device 1 flashes red is still inferred from the mean value of ${\sigma_x}^1$.

There are six spin observables involved in our story: ${\sigma_x}^1$, ${\sigma_y}^1$, ${\sigma_x}^2$, ${\sigma_y}^2$, ${\sigma_x}^3$, ${\sigma_y}^3$. The new feature of this example (called the GHZ example) is that we can also construct new observables by considering products and sums of these six observables.

It is an important fact in QM that the product of two observables is itself an observable if and only if the two observables commute.[8] It is therefore important to note that operators pertaining to different particles always commute. For example, ${\sigma_x}^1$ commutes with ${\sigma_y}^2$, and the product observable ${\sigma_y}^1{\sigma_x}^2$ commutes with ${\sigma_y}^3$, and so on. By using these facts alone, it follows that every product such as ${\sigma_y}^1{\sigma_x}^2{\sigma_y}^3$ is a QM observable. The product observables that correspond the three random variable products that appear in (1) are $\left({\sigma_x}^1{\sigma_y}^2{\sigma_y}^3\right)$, $\left({\sigma_y}^1{\sigma_x}^2{\sigma_y}^3\right)$, and $\left({\sigma_y}^1{\sigma_y}^2{\sigma_x}^3\right)$. Each observable has well defined mean values and variances in every quantum state.

It is interesting to note that these observables are mutually incompatible in the sense that one and only one can be measured. For example, $\left({\sigma_x}^1{\sigma_y}^2{\sigma_y}^3\right)$ is measured if and only if the devices are in the x-y-y setting. Yet, a simple calculation shows that any two of these products commute because we end up applying the anti-commutation relation twice. Therefore, there exists a quantum state that is an eigenstate of all three observables simultaneously. If the system is in this state, then the outcome of each product observable can be predicted with certainty. But none of the individual observables can be predicted with certainty. In fact, individual spin outcomes have probability ½ in this state (see Appendix). So there are examples in QM in which correlations are predicted with certainty, but the correlata are completely random.

The quantum mechanical story begins with the assumption that all the electron triples are prepared in the same quantum state, $|\psi\rangle$. The fact that there is always an odd number of red flashes in the settings y-y-x, y-x-y, and x-y-y tells us that the product observables are dispersion-free in the state $|\psi\rangle$, which implies that $|\psi\rangle$ is an eigenstate of the three product observables. In other words, the first set of experimental facts is accommodated by the supposition that $|\psi\rangle$ is in a $+1$ eigenstate of all of the observables $\left({\sigma_x}^1{\sigma_y}^2{\sigma_y}^3\right)$, $\left({\sigma_y}^1{\sigma_x}^2{\sigma_y}^3\right)$, and $\left({\sigma_y}^1{\sigma_y}^2{\sigma_x}^3\right)$. Therefore (1) is replaced by the laws:

$$\sigma_x^1\sigma_y^2\sigma_y^3\,|\psi\rangle = |\psi\rangle\,, \sigma_y^1\sigma_x^2\sigma_y^3\,|\psi\rangle = |\psi\rangle\,,\ \text{and}\ \sigma_y^1\sigma_y^2\sigma_x^3\,|\psi\rangle\,. \qquad (2)$$

What predictions can be made from these quantum mechanical constraints? The fact that any spin operator times itself is equal to the identity operator (as can be verified directly by squaring the matrices derived earlier in this section) and the fact that operators pertaining to different particles commute, proves that

$$\begin{aligned}(\sigma_x{}^1\sigma_y{}^2\sigma_y{}^3)(\sigma_y{}^1\sigma_x{}^2\sigma_y{}^3)(\sigma_y{}^1\sigma_y{}^2\sigma_x{}^3) &= (\sigma_x{}^1\sigma_y{}^2\sigma_y{}^3)(\sigma_y{}^3\sigma_x{}^2\sigma_y{}^1)(\sigma_y{}^1\sigma_y{}^2\sigma_x{}^3)\\ &= \sigma_x{}^1\sigma_y{}^2\sigma_x{}^2\sigma_y{}^2\sigma_x{}^3.\end{aligned}$$

Furthermore, the anti-commutation property of the spin operators proves that

$$\sigma_x{}^1\sigma_y{}^2\sigma_x{}^2\sigma_y{}^2\sigma_x{}^3 = \sigma_x{}^1(-\sigma_x{}^2\sigma_y{}^2)\sigma_y{}^2\sigma_x{}^3 = -\sigma_x{}^1\sigma_x{}^2(\sigma_y{}^2\sigma_y{}^2)\sigma_x{}^3 = -\sigma_x{}^1\sigma_x{}^2\sigma_x{}^3,$$

where the minus sign arises from the anti-commutation of $\sigma_x{}^2$ and $\sigma_y{}^2$. Recall that the anti-commutation law is required in order to accommodate the facts about sequential measurements (Section on Spin Measurements on a Single Electron).

Finally, from (2), it is also true that

$$(\sigma_x{}^1\sigma_y{}^2\sigma_y{}^3)(\sigma_y{}^1\sigma_x{}^2\sigma_y{}^3)(\sigma_y{}^1\sigma_y{}^2\sigma_x{}^3)\,|\psi\rangle = |\psi\rangle\,, \tag{3}$$

and therefore,

$$\sigma_x{}^1\sigma_x{}^2\sigma_x{}^3\,|\psi\rangle = -\,|\psi\rangle\,. \tag{4}$$

This proves what I promised: The *only* way for the quantum model to accommodate the first set of experimental facts is to assume that (2) is true, which implies (4), which then implies that there will always be an *even* number of red flashes when the magnets are set to x-x-x. This is the prediction that the hidden variable theory got wrong.

It is interesting to compare the hidden variable prediction with the QM prediction in this example. The hidden variable deduction proceeds in a completely analogous way up to the step where the anti-commutation relation is used. That step is replaced by the commutation relation $Y_2X_2 = X_2Y_2$, which then leads to a plus sign instead of the minus sign. All variables commute, while operators may or may not. In this respect, QM is more versatile. But don't mistake flexibility for weakness; for the flexibility allows us to embed all the spin 1/2 states in a single 2-dimensional vector space, and the tightness of the colligation leads to precise predictions about a variety of phenomena.

It is equally interesting to note the *similarities*. The predictive power of both theories relies on the idea that the variables or observables have values even when they are not measured. In the hidden variable prediction, this assumption comes into play when we assume that $Y_1{}^2 = Y_2{}^2 = Y_3{}^2 = 1$. For this to make sense, we assume that each variable has a value, and then argue that the square is 1 no matter which value it has. In the QM derivation, we assume that $\sigma_y^1\sigma_y^1 = \sigma_y^2\sigma_y^2 = \sigma_y^3\sigma_y^3 = I$, where I is the unit operator. Every state is an eigenstate of the unit operator, so this observable has the value 1 in every state. We are not assuming that σ_y^1, σ_y^2, and σ_y^3 have values in the sense of having dispersion-free probability distributions. But we are assuming that σ_y^1, σ_y^2, and σ_y^3 apply to situations in which they are not measured.

In fact, spin observables are extraordinarily projectible. For the state $|\psi\rangle$ is *uniquely* determined by the equations (2) *alone* (modulo a complex phase factor, which makes no difference to any prediction). This unique state $|\psi\rangle$ then determines the full probability distributions for *any* measurement settings whatsoever, including cases in which the three magnets are aligned in directions different from x and y. This is worth proving in detail because it illustrates the predictive power of the quantum mechanical colligation (see the Appendix). In particular, it is surprising to see that the unified QM model *requires* that each individual event is random (has probability $1/2$). QM isn't forced to explain why two people meet at the market by explaining why each will be there at that time. In comparison, the local hidden variable theory has very little predictive power; and to add insult to injury, the one prediction that it does make turns out to be false!

It is not my purpose to disparage hidden variable theories in general. After all, the unknown microstate of a thermodynamic system plays the role of a hidden variable in statistical mechanics. Rather, the point is that the hidden variable model does not compete with the unity of the QM model. The existence of hidden variables does not lead to very much predictive power, even though many philosophers consider them to be highly *explanatory*. And at the same time, QM is considered to be mysterious and un-explanatory despite the fact that it is predictively very powerful. So much the worse for explanation. Unification is a far better criterion of success.

After working through an example like this, after seeing how tightly quantum mechanics ties the phenomena together, wouldn't it be miraculous if this huge body of data were to fit the predictions of quantum mechanics without there being some way of explaining this fact in terms of a reality behind the observed phenomena? Indeed, the argument for a realist view of the QM properties of spin looks similar to the argument for the existence of Newtonian mass. In both cases, there is an empirical overdetermination of the values and properties of the postulated entities. There are many voices, and the theory predicts that they will sing in harmony. Hidden variable theory doesn't hear these voices, and what it does hear is not very harmonious.

Are Quantum Phenomena Explainable?

There is a wrong way and a right way of thinking about Reichenbach's cubical world. The wrong way is to view it in terms of Reichenbach's principle of common cause, which may be made more precise in the following way. For any two physical quantities X and Y that exhibit a statistical correlation, such that the correlation does not arise from X directly causing Y or Y directly causing X. Then, the only alternative appears to be that the correlation is explained by a common cause variable Z, such that Z causes X and Z causes Y.[9] This very powerful idea is severely challenged by QM examples.

A slight modification of the GHZ example brings the challenge into focus. Suppose that one measurement device is placed in Alice's laboratory, while the other two are placed in Bob's laboratory. The magnets are all aligned in the x-direction, so there is always an even number of red flashes. Bob hooks up a master light bulb that flashes red if and only one of his devices flashes red and the other flashes green. This means that Alice's light flashes red if and only if Bob's master light flashes red. There is a perfect positive correlation between what Bob sees and what Alice sees. Moreover, if we imagine that Alice and Bob measurements are performed simultaneously (in their common inertial frame of reference), then the theory of relativity appears to rule out the existence of any direct causal interaction between the events. The causal explanation recommended by Reichenbach's principle of common cause is ruled out by the Bell argument (Section on The Bell Theorem as Proving a Failed Consilience).

The Bell argument only rules out *local* hidden variable explanations and therefore only *local* causal models. Bell's argument is not a proof that QM is complete; it only proves that the EPR argument for the incompleteness of QM is unsound. If we allow some kind of action-at-a-distance, or backwards causation, then it is possible to avoid the false predictions. But avoiding false predictions is not the same as making the correct predictions. A causal model must replicate the predictions of QM and then make predictions that QM is presently unable to make. While there is no proof that it can't be done, there is no indication at the present time that this is a fruitful research strategy.

There are two desiderata that should be met if some kind of causal model is to explain the QM phenomena. The first fact to be explained is the remarkable connection between spin statistics in QM and facts about the geometry of space and time. In Chapter 6 of Feynman et al. (1965), Feynman takes the argument one step further. Instead of merely showing how the properties of entangled pairs and triplets of particles follow from the properties of spin required to accommodate the statistics exhibited in single electron experiments, he shows how the single electron spin statistics follow from the isotropy of space. Recall from the section on Spin Measurements on a Single Electron that the probability of getting spin up in direction v given that the electron in spin up in direction u is equal to $\cos^2(\theta/2)$, where θ is the angle between u and v. The fact that the conditional probabilities are a function of θ only shows that they are same if the whole apparatus is rotated in space. The directions u and v will change, but the angle between them will be the same. This shows that the QM probabilities are invariant under space rotations, as required. But there are many ways of satisfying this requirement. In fact, so long as the conditional probabilities are some function of θ, the isotropy of space is preserved. What Feynman proves is that the particular function $\cos^2(\theta/2)$ follows *uniquely* from the isotropy of space plus the basic principles of QM. Those basic principles state merely that spin $1/2$ states are represented vectors in a 2-dimensional complex vector space (a Hilbert space) with the usual connection states and probabilities (the Born rule). This shows that there is a broader consilience between QM and the geometry of space, and such consiliences must be maintained by any superseding theory.

The second desideratum is that there should be some explanation of why QM uses operators in place of variables. One difference is that variables always have definite values, whereas operators only have definite values (dispersion free values) when the particle is in an eigenstate of the operator. In all other cases, the theory assigns only a probability distribution. The assertion that the observable has a value if and only if it is an eigenstate of the corresponding operator is commonly called the eigenstate–eigenvalue link; it is taken seriously in some interpretations of QM (the only-if part is denied by hidden variable theories). A dramatic example of what it means to take the eigenstate–eigenvalue link seriously occurs in the case of the position observable. If a wave function is not in a extremely spiked Dirac delta state, then it's not in an eigenstate of position, and therefore has no position. Therefore, any particle with that is "spread" over an extended region of space is not in space. So, whatever causal stories are told about quantum particles, they should allow that causal processes take place outside of space, and possibly outside of time.

Measurements on this view are also more complicated kinds of causal processes. If a system has no precise value of an observable before it is measured, and the system ends up with a definite value (the measured value), then the measurement must have disturbed the state of the system. This is a collapse theory of QM measurement. It is contrasted with the view that measurements merely allow particles in some states to pass through the device and block others, without changing the states of the particles selected – this is a hidden variable view because it assumes that the particles have properties to be filtered. But as Bell has shown, even this view is committed to some kind of measurement disturbance by which non-local events sometimes change the states of distant particles. It is hard to escape the conclusion that QM measurement is more complicated than is assumed in classical physics.

A fundamental requirement is that any theory that supersedes current theories in physics should *preserve the consilience of inductions* already achieved by QM. Without that condition, the new theory falls short. It's not enough to merely avoid false predictions. A new theory must reproduce the predictions of the old theories, at least to the degree to which they have proved to be accurate. The "old" theories of physics include the quantum theory and the theory of relativity, and quantum field theory. It is a problem that requires considerable invention and innovation. A metaphysics of causal processes involving "events" that take place outside of space and time is not easy and obvious.

One approach has been developed by Alexey Kryukov (2003, 2004, 2005, 2006, 2007). The idea is that the ordinary three-dimensional space of space of our everyday experience is actually a three-dimensional sub-manifold of points in a larger infinite-dimensional space, which is defined quantum mechanically. Space is not postulated as a pre-existing structure over which wave functions are defined. Rather, on this quantum geometrical view, the metric properties of space are *derived* from the metric properties of the infinite-dimensional space in which it is embedded. Causal processes and 'events' are located in the enveloping "hyperspace". Kryukov (2008) has extended this theory to the Minkowski spacetime of special relativity and Kryukov (forthcoming) proves that the curved spacetime of general relativity can also be derived from QM structures. It is a severe constraint to require that the metric structure of spacetime be *determined* from the metric structure of a QM state space. In this case, it requires that the QM state space is suitably modified so that "squared distances" can be negative or positive, or zero. If a very severe constraint is able to reproduce the consiliences of older theories in a unified way, then it also has the potential to lead to new consiliences and novel predictions. Time will tell.

To sum up: The right way of thinking about Reichenbach's cubical world is in terms of the consilience of inductions. There is no categorical imperative that we must represent the external world in terms of variables as opposed to the more geometrical concepts used in QM. If we go by the consilience of inductions, then local hidden variable theories do not succeed and non-local hidden variable theories currently fail to compete with QM. The miraculous consilience of QM not only explains the growing consensus in favor of QM, but it also explains why cubical world inference is so hard to characterize in general terms. For the inference that is ultimately convincing most often depends on the explication of new theoretical concepts and new mathematical constructions.

The good news is that there do appear to be methodological principles that apply equally validly to the old and the new physics. As in science itself, successful predictions in philosophy are convincing evidence in favor of a theory. In 1858, Whewell's concluded that the consilience of inductions is the best indicator of good science. The theories have changed, but Whewell's prediction appears to be right.

Appendix

The purpose of this appendix is to prove that the quantum mechanical accommodation of the first three experimental facts in the GHZ experiment leads to far stronger predictions than those provided by the hidden variable theory.

Because there are eight possible outcomes, the state vector $|\psi\rangle$ is a vector in an eight-dimensional Hilbert space. This space is constructed out of the three two-dimensional vector spaces used to represent the states of electrons separately. Let us choose the basis vectors for the first Hilbert space to be $\left|z^{+}\right\rangle_1$ and $|z^{-}\rangle_1$, etc., where the subscripts keep track of the vector space in question. It is now possible to prove that the eight vector products, $\left|z^{\pm}\right\rangle_1 \left|z^{\pm}\right\rangle_2 \left|z^{\pm}\right\rangle_3$, form a natural basis for the eight-dimensional (complex) vector space. It is convenient to denote these eight basis vectors by $|\pm\pm\pm\rangle$.

From the section The Consilience of Quantum Mechanics, σ_x has the property

$$\sigma_x \left|z^{+}\right\rangle = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = |z^{-}\rangle .$$

Similarly, $\sigma_x |z^{-}\rangle = \left|z^{+}\right\rangle$. In the context of the three-electron system, a spin operator like ${\sigma_x}^3$ operates on the third part of the vector product, so for example,

$$\sigma_x^3 |+++\rangle = |++-\rangle .$$

Also,

$$\sigma_y \left|z^{+}\right\rangle = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = i \begin{bmatrix} 0 \\ 1 \end{bmatrix} = i\, |z^{-}\rangle .$$

Similarly, $\sigma_y \left|z^-\right\rangle = -i \left|z^+\right\rangle$. We now have everything in place that we need to apply the defining GHZ equations (2) to an *arbitrary* state vector

$$|\psi\rangle = \sum_{j,k,l} c_{jkl} \, |jkl\rangle,$$

where *j, k*, and *l*, range over the values $+1$ and -1. Calculating the constraints on the coefficients c_{jkl} implied by (2) is tedious, but it involves nothing more than algebra. The key point is that an operator, such as $\sigma_x^1\sigma_x^2\sigma_x^3$ induces a one-to-one map from basis states to basis states. Thus, for example, $\sigma_x^1\sigma_x^2\sigma_x^3 \, |+++\rangle = |---\rangle$. So from $\sigma_x^1\sigma_x^2\sigma_x^3 \, |\psi\rangle = -\, |\psi\rangle$, it follows that $\sigma_x^1\sigma_x^2\sigma_x^3 c_{+++} \, |+++\rangle = c_{+++} \, |---\rangle = -c_{---} \, |---\rangle$, and so on, for each of the eight components. Therefore, $c_{+++} = -c_{---}$, and more generally, $c_{jkl} = -c_{-j-k-l}$, or $c_{\pm\pm\pm} = -c_{\mp\mp\mp}$. The requirement that $\sigma_x^1\sigma_y^2\sigma_y^3 \, |\psi\rangle = |\psi\rangle$ is going to agree with the equation just derived because $\sigma_x^1\sigma_y^2\sigma_y^3 c_{\pm++} \, |\pm++\rangle = i^2 c_{\pm++} \, |\mp--\rangle = c_{\mp--} \, |\mp--\rangle$, but in the other cases, $\sigma_x^1\sigma_y^2\sigma_y^3 c_{\pm-+} \, |\pm-+\rangle = -i^2 c_{\pm-+} \, |\mp+-\rangle = c_{\mp+-} \, |\mp+-\rangle$, it disagrees in the sign, thereby implying that $c_{\pm-+} = c_{\mp+-}$. Since we have already shown that $c_{\pm-+} = -c_{\mp+-}$, the only solution is that $c_{\pm-+} = c_{\mp+-} = 0$. The rule is that if the two signs for the y values disagree (one is $+$ and the other is $-$), then the coefficient is 0. By applying the same rule to the other two equations, $\sigma_y^1\sigma_x^2\sigma_y^3 \, |\psi\rangle = |\psi\rangle$, and $\sigma_y^1\sigma_y^2\sigma_x^3 \, |\psi\rangle = |\psi\rangle$, we find that the only no-zero coefficients are c_{+++} and c_{---}, and $c_{+++} = -c_{---}$. Therefore, the unique GHZ state is, modulo an arbitrary phase factor $e^{i\theta}$,

$$|\psi\rangle = \tfrac{1}{\sqrt{2}} \, |+++\rangle - \tfrac{1}{\sqrt{2}} \, |---\rangle \, .$$

It is especially important to notice that any three of the equations is sufficient to determine a unique state vector and predict the remaining equation.

Notes

1 This mirror example is used in Hung (1997).

2 The principle is attributed to William of Ockham ($\sim$1280–1347 AD), although it has since taken on a life of its own.

3 It is no coincidence that Newton scholars, such as Harper (2002), emphasize the importance of the agreement of independent measurements in Newton's argument for universal gravitation. For Whewell was also primarily concerned with the explication of Newton's methodology. See also Myrvold and William (2002) for an argument that this kind of evidence is not properly taken into account in standard statistical methods of model selection.

4 My purpose is not to argue for some particular historical or exegetical thesis about Whewell's notion of consilience, but to use (and adapt) Whewell's idea for the purpose of explaining how probabilistic theories work in general, and how quantum mechanics works in particular.

5 Recall that a random variable is, by definition, any variable that has a probability distribution associated with it.

[6] As far as I'm aware, this exact experiment has not been performed. My confidence in making this claim is based on the proven consilience of QM (next section).
[7] In the general case in which observables involve complex numbers, only Hermitian operators count as observables, because only they have expected values that are real numbers in every state.
[8] This is because the product of two Hermitian operators is Hermitian if and only they commute.
[9] There are a number of other scenarios that need to be ruled out as well. Arntzenius (1993) provides a concise list of these. Also see Sober (1984), Cartwright (1989), Eells (1991), Hausman (1998), and Woodward (2003) for extensive discussions of causal inference and explanation, and related philosophical issues.

References

Arntzenius F (1993) The common cause principle. PSA 1992, vol 2. Philosophy of Science Association, East Lansing, Michigan, pp 227–237

Arntzenius F (1997) Transition chances and causation. Pacific Philos Quart 78:149–168

Bell JS (1964) On the Einstein–Podolsky–Rosen paradox. Physics 1:195–200

Bell JS (1971) Introduction to the hidden variable question. In: d'Espagnat B (ed) Foundations of quantum mechanics. Academic, New York

Butts RE (ed) (1989) William Whewell: theory of scientific method. Hackett Publishing Company, Indianapolis/Cambridge

Cartwright N (1989) Nature's capacities and their measurement. Oxford University Press, Oxford

Eells E (1991) Probabilistic causality. Cambridge University Press, Cambridge

Einstein A, Podolsky B, Rosen N (1935) Can quantum-mechanical description of physical reality be considered complete? Phys Rev 47:777–780

Feynman R, Leighton, Sands (1965) The Feynman lectures on physics: quantum mechanics, vol III. Addison-Wesley, Reading, MA

Forster MR (1984) Probabilistic causality and the foundations of modern science. Ph.D. thesis, University of Western Ontario

Forster MR (1986) Bell's paradox and path analysis. In: Weingartner P, Dorn G (eds) Foundations of physics. Holder-Pichler-Tempsky, Vienna, pp 191–226

Greenberger DM, Horne MA, Zeilinger A (1989) Going beyond bell's theorem. In: Kafatos M (ed) Bell's theorem, quantum theory and conceptions of the universe. Kluwer, Dordrecht, pp 69–72

Harper WL (2002) Howard Stein on Isaac Newton: beyond hypotheses. In: David BM (ed) Reading natural philosophy: essays in the history and philosophy of science and mathematics. Open Court, Chicago/La Salle, IL, 71–112

Hausman DM (1998) Causal Asymmetries. Cambridge University Press, Cambridge

Hung E (1997) The nature of science: problems and perspectives. Wadsworth Publishing Co, Belmont, CA

Khinchin AI (1960) Mathematical foundations of quantum statistics. Dover, Mineola, NY

Kryukov A (2003) Coordinate formalism on abstract hilbert space: kinematics of a quantum measurement. Found Phys 33(3):407–443

Kryukov A (2004) On the problem of emergence of classical space–time: the quantum-mechanical approach. Found Phys 34(8):1225–1248

Kryukov A (2005) Linear algebra and differential geometry on abstract Hilbert space. Int J Math Math Sci 14:2241

Kryukov A (2006) Quantum mechanics on Hilbert manifolds: The principle of functional relativity. Found Phys 14:2241–2275

Kryukov A (2007) On the measurement problem for a two-level quantum system. Found Phys 37:3–39

Kryukov A (2008) Nine theorems on the unification of relativity and quantum mechanics. J Math Phys 49:102–108

Kryukov A (2010) A possible mathematics for the unification of quantum mechanics and general relativity. J Math Phys 51:022110

Mermin DN (1990) Quantum mysteries revisited. Am J Phys 58:731–734

Myrvold W, William LH (2002) Model selection, simplicity, and scientific inference. Philos Sci 69:S135–S149

Reichenbach H (1938) Experience and prediction. University of Chicago Press, Chicago, IL

Reichenbach H (1956) The direction of time. University of California Press, Berkeley, CA

Sober E (1984) The nature of selection: evolutionary theory in philosophical focus. MIT Press, Cambridge, MA

Sober E (1994) Temporally Oriented Laws in Sober (1994). From a biological point of view – essays in evolutionary philosophy. Cambridge University Press, Cambridge, pp 233–251

van Fraassen BC (1982) The Charybdis of realism: epistemological foundations of Bell's inequality argument. Synthese 52:25–38

Whewell W (1858) Novum Organon Renovatum, Part II, 3rd edn. The philosophy of the inductive sciences. London, Cass, 1967

Woodward J (2003) Making things happen: a theory of causal explanation. Oxford University Press, Oxford/New York

Probability and Objectivity in Deterministic and Indeterministic Situations*

James H. Fetzer

The subject I would like to address, namely, probability and objectivity in deterministic and indeterministic situations, could be formulated by means of the question, "To what extent does the propensity approach to probability contribute to plausible solutions to various anomalies which occur in quantum physics?" In order to pursue this problem, I shall, first, sketch several of these anomalous conditions, second, clarify the difference between deterministic and indeterministic situations, and, third, consider three alternative interpretations of probability as they apply to these problems, with particular reference to Einstein's criterion of physical reality. The position I shall defend is that of these three interpretations – the frequency, the subjective, and the propensity – only the third accommodates the possibility, in principle, of providing a realistic interpretation of ontic indeterminism. If these considerations are correct, therefore, they lend support to Popper's contention that the propensity interpretation tends to remove (at least some of) the mystery from quantum phenomena (Popper 1957, 1967, 1982).

J.H. Fetzer (✉)
Division of Humanities, New College of the University of South Florida Sarasota, FL 33580, USA
e-mail: jfetzer@d.umn.edu

*This paper is an expanded and revised version of a lecture presented at the University of Colorado in February and at New College in April 1983. I am grateful to Paul Humphreys and to Stephen Brush for their comments and inquiries. In particular, Professor Brush has raised the issue of the Aspect experiments – which *seem* to confirm action at a distance – in relation to the approach defended here. As Popper (1982, p. xviii) has noted, this result need not undermine realism; but it would be remarkable if quantum mechanics could be true only if relativity were false. If situations involve propensities if and only if they are indeterministic in character, as my position implies, then if experiments such as Aspect's, Einstein-Podolsky-and-Rosen's, and others are indeterministic, then propensities are involved. If they are *both* indeterministic *and* action at a distance cannot be avoided (by appealing to non-causal, but not therefore non-lawful, relations, for example), which I doubt, then indeterministic causal concepts as well as deterministic causal concepts will almost certainly require revision. In any case, while the propensity approach can contribute to the resolution of some of the anomalies arising within the quantum domain, others – some involving questions of completeness – no doubt remain.

E. Eells and J.H. Fetzer (eds.), *The Place of Probability in Science*, Boston Studies in the Philosophy of Science 284, DOI 10.1007/978-90-481-3615-5_10,

The Dimensions of the Problem

To offer some background for the sorts of issues that seem to be involved here, the sources of anomaly that occur in quantum physics arise from at least three apparently different types of situations. The first are characterizable by means of Heisenberg's Uncertainty Principle, according to which the measure of certainty or of precision with which the position p and the momentum q of a sub-atomic particle may be specified are inversely related, such that the more precisely its momentum can be measured, the less precisely its position, where there is a limit with respect to simultaneous measurements of position and of momentum reflected by the Uncertainty Principle itself, namely:

(1) $\Delta p \Delta q \geq h$

where the product of Δp and of Δq as measures of uncertainty of position and of momentum, respectively, can never violate the limiting value of h, understood as Planck's constant, a very small unit of energy.

Now this notion of uncertainty is amenable to at least two kinds of interpretation. One interpretation is an "ontic" interpretation, which would have us understand that sub-atomic particles simply do not *have* simultaneous position and momentum, i.e., that we ought to read formula (1) as a property of the physical world, independently of our knowledge or belief or awareness thereof. The other is an "epistemic" interpretation, which instead asserts that the uncertainty relations are properly regarded as representing limitations upon how much we can *know* about the simultaneous position and momentum of sub-atomic particles. Popper has suggested that the ontic interpretation is farfetched, especially since determinations of momentum are usually made by means of position measurements at different times, say, t_1 and t_2, rather than by means of a simultaneous determination (Popper 1967). An arrangement consisting of two photographic plates separated by a fixed distance, say, d, could be employed to measure the location of a particle at time t_1 and later at time t_2, so that by comparing the results we can compute not only where a particle was at time t_1 but also where it was going.

The determination of momentum by this means, however, is subject to certain qualifications; for the magnitude obtained by dividing the distance, d, by the time, $t_2 - t_1$, may represent a *mean* (or "average") value – unless there are grounds to assume that particles always travel with constant momentum, a doubtful contention. Moreover, if the distance, d, were made less and less, it might still remain the case that there is a limit on the precision attainable as a function of natural law, such that formula (1) reflects an irreducible *ontic* feature of the physical world, after all. Still, I am inclined to agree with Popper that the denial that particles have simultaneous position and momentum is indeed far-fetched, even if the Uncertainty Principle does reflect a limitation, in principle, upon what we can *know*, i.e., as a function of natural law. But it would be valuable to discover an adequate explanation for this perplexing situation – perhaps, for example, by arriving at a theoretical understanding of the second source of anomaly which I would like to consider, the phenomenon known as "wave-particle duality".

Wave-particle duality may be illustrated by (what are usually described as) "one slit" experiments, in which a source of electrons, A, let us say, is directed through a single aperture toward an apparatus, such as a photographic plate, once again, which is capable of detecting their arrival. What has been discovered is that the behavior of electrons is such that, when this experiment is subject to replication, it is not the case that they always wind up at just one place or location or area; instead, we obtain something like a distribution of impact, such that, if various areas of impact were labeled "B", "C", and so on, then with a certain probability, represented by the absolute value of the appropriate psi-function squared, $|\Psi_B|^2$, $|\Psi_C|^2$, and so forth, an electron will end up in one area, say, B, rather than another, C. The principal problem that arises here is what we are supposed to understand by the term "probability" as it is employed within this context. But notice that, even if we assume that particles always have both position and momentum during every moment of their history, nevertheless, the distribution of their impact – together with the interference patterns displayed during other, "two slit", experiments – represents wave-like behavior by particle-like entities, which appears to be very difficult to comprehend.

The third source of anomaly I want to mention concerns the point of view known as "complementarity", which was originally introduced by Bohr to designate the necessity to depend upon two (seemingly inconsistent) conceptions of physical reality, i.e., the wave picture and the particle picture, without assuming any prospects for their theoretical reconciliation. "Complementarity", of course, has come to mean many things to many people, not least of whom are those allied with the Copenhagen Interpretation of quantum physics, according to which the human mind performs an essential role in understanding the physical world, not only in the acquisition of information about the world but in bringing about the existence of properties of the world by means of causal interaction through the process of measurement. Thus, Einstein tended to differ with Bohr with respect to the possible existence of unmeasured properties, rather than following Bohr in maintaining that,

> a quantity is not real just because it *can* be measured, it is also necessary that it *is* measured. Or rather, reality cannot be attributed to the quantity itself in any case, but only to the measurement of the quantity. (Brush 1979, p.93; for alternative interpretations, see Folse 1977)

While it seems quite plausible to suppose that *measurements* arise as a result of a causal interaction between a *measuring instrument*, i.e., a device capable of making measurements, and a *measureable property*, i.e., a property which is capable of being measured, it does appear anomalous to suppose that properties themselves only exist so long as they are being subjected to measurement.

This situation is further complicated, of course, by the introduction of probabilistic properties, whose measurement involves problems not encountered within other contexts. An example sometimes offered to illustrate the difficulties confronting the Copenhagen Construction is that of Schrödinger's "Cat Paradox", which envisions an arrangement consisting of a living feline placed inside an opaque chamber that is connected to an electrical device activated by a radioactive source linked to a geiger counter. If decay occurs and registers – with a probability, let us assume, equal to 1/2 – then an impulse is activated which electrocutes the cat. From the

Copenhagen Perspective, it seems, until the chamber has been opened and a live kitty or a dead corpse has been observed, the cat itself is presumed to be neither dead nor alive but rather somewhere in between, which is represented by a superposition of psi-functions for both of these events prior to this measurement being made. The procedure of taking a look and discovering, say, that the cat is dead thus effectuates (what is usually referred to as) "the reduction of the wave packet", whereby, in effect, the result of being dead or of being alive is brought about. All of this is anomalous, indeed, because we are inclined to believe that the cat, after all, is either dead or alive during each moment of its history, whether we happen to notice or not. So if there is some way around problems such as these, then it would appear very desirable to pursue, where much of what follows is intended to suggest that the Copenhagen Interpretation is not only an implausible approach toward quantum physics but also an avoidable one.

Interpreting "Indeterminism"

In order to answer the question, "Can there be a realistic interpretation of ontic indeterminism?", it is indispensable to clarify the special character of ontic indeterminism, on the one hand, and of realistic interpretations, on the other. What I want to do is to consider the criterion of reality advanced by Einstein in the famous paper he co-authored with Podolsky and Rosen – where they attempt to define conditions under which the existence of an element of physical reality should be inferred – to ascertain, in part, whether different conceptions of probability can make a difference to our understanding of what ontic indeterminism itself entails. But before going that far, let us observe that three different types of situations appear to require differentiation as follows, namely:

(2)

Three situation are possible:	Ontic determinism	Epistemic determinism
a. Classical Mechanics (CM):	Yes	Yes
b. Statistical Mechanics (SM):	Yes	No
c. Quantum Mechanics (QM):	No	No

With respect to each of these areas of inquiry, the issue arises, "Does the domain in question – classical, statistical, and quantum phenomena, respectively – represent ontic determinism/indeterminism or does it instead reflect merely epistemic determinism/indeterminism?" Indeed, these distinctions apply to theories as well as to phenomena in each of these domains. Classical mechanics, for example, seems to represent ontically deterministic phenomena where, if only the relevant parameters (or "initial conditions") are specified with sufficient precision, then it is possible, in principle, to predict the outcome that will occur in every single case without exception, assuming, of course, some such theory happens to be true. Loosely – intuitively – let us suppose that this conception affords an appropriate foundation for understanding ontic determinism, i.e., that the same outcomes are always brought about under the same initial conditions, where exact predictions, in principle, are always possible, relative to true theories for those domains.

In classical mechanics, moreover, we are also in a situation where quite often these initial conditions are ascertainable, in practice, which makes it possible to successfully predict. Thus, under these conditions, epistemic as well as ontic determinism obtains. In the case of statistical mechanics, however, the situation appears to be somewhat different, not with respect to ontic determinism but rather with respect to epistemic determinism. For, here we seem to be dealing with circumstances – as in the case of gas laws – involving enormous numbers of very small things, where, even if we take for granted that, in principle, these enormous numbers of very small things behave in the small exactly the way billiard balls, for example, behave in the large, nevertheless, they are so many and so small that, as a practical matter, it is virtually impossible for us to possess the kind of knowledge about their initial conditions that would be required for us to be successful in predicting their behavior – even though we take it to be the case that statistical mechanistic phenomena are ultimately ontically deterministic, i.e., such that, if only we *could* have sufficiently precise measurements of these initial conditions, the appropriate laws of statistical mechanics *could* be applied and would yield an exact prediction of what would occur in every single case – without exception! Nevertheless, characteristically we are not in such a position, but rather in one of epistemic indeterminism, which can be viewed in several different ways where the phenomena are complex, where we have a lack of knowledge, and where statistical predictions are relied upon as a function of our ignorance.

In the case of quantum mechanics, by contrast, we seem to be in a rather different situation. Here it at least appears to be the case that we are possessed of enough knowledge about initial conditions that it is implausible to suppose that the situation is merely one of epistemic indeterminism. In fact, no matter how hard we look, how much we try, how many guesses we explore, it is very difficult to come up with any additional factors which make a difference, for example, to an electron's probability of landing in area B, or area C, and so on, upon its emission from the source, A. This seems to be an irreducibly probabilistic or ontically indeterministic phenomenon, where different outcomes are sometimes brought about under the same initial conditions and exact predictions, in principle, are by no means always possible. To more adequately clarify the situation encountered, let us consider the following set of five alternative hypotheses which might be thought to be applicable here:

(3) Possible alternative hypotheses:

h_1: the deterministic hypothesis
h_2: the random-walk hypothesis

h_3: the free-will hypothesis
h_4: the probabilistic hypothesis
h_5: the hidden-variable hypothesis

Thus, according to h_1, the deterministic hypothesis, the relation between an electron's emission from source A and its impacting in area B, for example, is one of ontic determinism, where in every single case the same outcome occurs under the same initial conditions and exact predictions are always possible. However, if all of the known factors *are* a complete set of factors, then the deterministic hypothesis is just not true, since the available relevant evidence overwhelmingly supports the conclusion that the situation is not one of ontic determinism. So let us assume not-h_1.

According to h_2, the random-walk hypothesis, the way in which a particle gets somewhere from wherever it may be is in a wholly random way, because the particle, roughly speaking, has an equal probability (or a "classical" tendency) for travelling in every direction. The types of distributions obtained for impacts within specified areas such as *B, C*, and so forth, after emission from a source *A*, however, are not the types that would lend support to a random-walk hypothesis. So let us also assume not-h_2. According to h_3, the free-will hypothesis, of course, each electron simply "makes up its own mind" with regard to its own destination, as though electrons had minds to "make up". In the case of the free-will hypothesis, we in effect abandon predictability, because free-will, in this sense, really entails that such a phenomenon is not governed by natural laws but instead either lies beyond their scope or occurs in violation thereof. The adoption of this hypothesis would thus reflect the belief that quantum phenomena lie beyond the pale of science, which seems to be a logical possibility without adequate evidential warrant. For electrons do not behave as if they were "making up their own minds" but rather conform to probabilistic expectations, where these expectations appear to vary with variations in initial conditions. While we cannot predict where each single electron is going to impact, we can make statistical predictions of patterns of impact that will tend to be displayed by the impact of many electrons as a function of those initial conditions, which undermines a free-will hypothesis. So let us further assume not-h_3.

We seem to be left with two alternatives. According to h_4, the probabilistic hypothesis, the situation confronted here really is one of irreducible ontic indeterminism, where one or another of various different outcomes will be brought about under the same initial conditions, where exact predictions, as opposed to probabilistic expectations, are, in principle, not possible. This, of course, is a very tempting hypothesis, given the available evidence. Yet it is still not our only option, since according to h_5, the hidden-variable hypothesis, there may remain other factors, say, F^1, F^2, and so on, such that if these factors are included in a specification of initial conditions along with A, for example, then it might turn out to be the case after all that the relation between an electron's emission from a source, A, and its impacting in area B really is one of ontic determinism, relative to a complete specification of initial conditions $A \cdot F^1$, $A \cdot F^2$, and so forth. Remember that the original deterministic hypothesis, h_1, was predicated on the assumption that all of the *known* factors were all of the *relevant* factors in relation to this class of outcome phenomena, an assumption which the hidden-variable hypothesis, h_5, thus denies.

We have already discovered that the relevant evidence tends to falsify the deterministic hypothesis, h_1, but the hidden-variable hypothesis, h_5, is somewhat more subtle, insofar as it is based upon the following observation: even if we have looked very long and very hard and have not found any additional relevant factors, this in itself does not establish that they do not exist; it shows, on the contrary, that *either* they do not exist *or else* we have not yet looked in the right places to find them! Thus, someone who is very strongly committed to a deterministic thesis about the world could still preserve their commitment to determinism in the face of recalcitrant quantum phenomena by holding out for the existence of hidden variables, as hypothesis h_5 proposes, the discovery of which will eventually disclose the deterministic character thereof. Until very recently, moreover, it was very difficult to see how the hidden-variable hypothesis could be disconfirmed as well as confirmed; but the derivation of (what is known as) "Bell's inequality" has dramatically altered this situation, where Bell's inequality represents a set of relations which must be satisfied if hidden-variable hypotheses are true. When subjected to experiments (which are continuing today), the evidence has tended to falsify Bell's inequality – and with it the hidden-variable hypothesis, h_5. So let us tentatively assume not-h_5 as well.

Einstein's Reality Criterion

These considerations, if correct, suggest that ontic indeterminism may be theoretically unavoidable, especially if the probabilistic hypothesis h_4 happens to be true; but it sheds no light on whether or not a realistic interpretation of ontic indeterminism, in principle, might be possible at all. In order to pursue this issue, therefore, let us consider the criterion of reality advanced by Einstein in the context of his paper with Podolsky and Rosen, as follows:

(4) If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to that physical quantity. (Einstein et al. 1935, p. 777)

Certain infelicities, it must be admitted, attend this formulation. One is that it does not adequately differentiate between phenomena and theories as representatives of phenomena; for surely predictions are conclusions of inferences from premises, where lawlike sentences and descriptions of initial conditions characteristically serve as evidential warrants. Thus, it seems, we should think of Einstein as regarding a theory as describing a system in the world, where ascriptions of initial conditions assign specific values v_i which permit the derivation of specific predictions of other values v_j from certain theoretical assumptions. But another difficulty is thereby made apparent, for this formulation likewise fails to separate the ontic from the epistemic: even well-entrenched theories are likely to yield accurate predictions, without therefore being guaranteed to be true. Phlogiston, Allan Franklin reminds me, did not acquire existence as an element of physical reality in spite of innumerable "accurate predictions" of gains and losses of this property based upon phlogiston theory.

As a criterion of belief, in other words, there may be much to be said in support of the thesis that, under the conditions specified, an evidential warrant supports the *belief* that there exists an element of physical reality corresponding to such physical quantities. As a criterion of truth, however, there is little to be said in support of the claim that, under the specified conditions, an evidential warrant guarantees the *truth* that there exists an element of physical reality corresponding to such physical quantities. And clearly an adequate conception of a standard of existence should not commit any realist to the existence of properties stipulated by any theory – unless that theory is true! In order for Einstein's criterion of reality to be adequate, therefore, it has to be qualified with the condition, "so long as the theory from which those predictions were derived is true". If this condition is satisfied, however, then the rest of Einstein's criterion does not require satisfaction, since corresponding elements of physical reality must then exist whether or not they are ever subject to systematic prediction. From an ontic point of view, without this condition, Einstein's criterion is false; but with this condition, it is trivial.

These apparent shortcomings notwithstanding, Einstein's criterion has exerted considerable influence upon discussions of realism in the past and in the present; for this reason, as well as others, it will serve as an appropriate background for our discussion of ontic indeterminism. Thus, the formulation endorsed by Einstein bears a striking resemblance to a certain construction advanced by Reichenbach as his account of strict causal laws, which we shall refer to as "Assumption D" (for "determinism"), as follows:

Assumption D: The statement that nature is governed by strict causal laws means that we can predict the future with a determinate probability and that we can push this probability as close to certainty as we want by using a sufficiently elaborate analysis of the phenomena under consideration. (Reichenbach 1944, pp. 2–3)

Indeed, the resemblance is even closer than it initially appears, insofar as Reichenbach, like Einstein, interprets "certainty" as a probability equal to unity. Moreover, as Clifford Hooker has observed, Einstein almost certainly adopted another assumption as well, which we shall refer to as "Assumption E" (for "Einstein"), as follows:

Assumption E: A complete description of a physical system S during some time interval T is one for which every attribute of S is precisely determined for each instant $t \in T$. (Hooker 1972, p. 71)

These assumptions are theoretically significant for understanding Einstein's position, since they jointly entail the following claim, namely:

(5) If we can predict every attribute of a system S for each instant $t \in T$ with certainty, then (a) our description of S is complete and (b) S is governed by strict causal laws;

which supports the possibility that alternative formulations may be required to capture a realistic interpretation of ontic indeterminism.

Perhaps the most obvious logical feature of Einstein's criterion (4) is also one of its most important aspects; for, as he himself remarked, it is intended as a sufficient condition, but not as a necessary condition, for the value of a predicted quantity v_j to correspond to an element of physical reality. Its 'If ..., then _____' conditional structure reflects that this criterion of reality, such as it may be, is only meant to represent one way in which the existence of an element of physical reality may be ascertained. Thus, the crucial issue before us may be expressed in the following manner:

(6) If, without in any way disturbing a system S, we can predict, not with certainty but with some probability less than 1, the value of a physical quantity, can there then not still exist some element of physical reality corresponding to that physical quantity?

where it is especially important, in light of *Assumption D* and *Assumption E*, that Einstein's criterion as a *sufficient condition* not be confounded with, let us say, a *sufficient definition*, in the sense of a weakest criterion of reality, which would be a necessary and sufficient condition of the broadest general kind. For a realistic interpretation of ontic indeterminism to qualify as a theoretical possibility, in principle, after all, surely it must be at least logically possible (a′) that our description of a system S could be complete even though (b′) the system S is not governed by strict causal laws. Otherwise, the combined force of *Assumption D* with *Assumption E* as features of Einstein's criterion of reality strongly suggests if not strictly entails that realistic interpretations of quantum phenomena are deterministic, necessarily, in which case the probabilistic hypothesis h_4 could not possibly be true.

Indeterminism and Realism

We have discovered a striking dilemma, insofar as the arguments of the second section, which are based upon experimental findings, support the conception that a realistic interpretation of ontic indeterminism is a theoretical necessity, while the arguments of the third section, which are based upon Einstein's criterion, suggest the conclusion that a realistic interpretation of ontic indeterminism is a theoretical impossibility. This dilemma, of course, is more apparent than real, once Einstein's criterion has been diagnosed as a sufficient rather than a necessary condition. Disclosing the inadequacy of a deterministic criterion, moreover, is not the same thing as uncovering an adequate indeterministic criterion, but it still might pave the way; for the identification of "certainty" with probability equal to unity raises the possibility that the identification of "uncertainty" with probability less than unity deserves to be explored. The theoretical significance of the probabilistic hypothesis, h_4, in relation to quantum phenomena, after all, cannot be ascertained without consideration for the interpretation of probability that effects its connection with this domain, where three interpretations appear to warrant special consideration within this specific context, namely: the frequency, the subjective, and the propensity conceptions, respectively.

According to the frequency conception, probabilities are limiting frequencies for specified attributes, such as impacting in area B, within designated reference classes, such as emissions of electrons from sources such as A. A probabilistic hypothesis under the frequency interpretation may be characterized as possessing the following logical form,

(7) $P(B/A) = p$

where 'p' denotes the limiting frequency for the attribute B within an infinite reference class A. If the probability with which an electron emitted from source A lands in area B is given by the absolute value of the corresponding psi-function squared, $|\Psi_B|^2 = \frac{1}{4}$, say, this means that the limiting frequency with which impacts in area B should be expected to occur over an endless sequence of emissions from source A is equal to $\frac{1}{4}$. If a reference class happens to be finite, 'p' may be taken as the limit that would obtain were its members repetitiously counted over and over again. Since values of limits are logically compatible with any number of exceptions n, of course, the identification of "certainties" (in Einstein's sense) with probabilities of one (or of zero) appears to be inappropriate; but this is not a difficult problem to repair, since constant conjunctions (and constant non-occurrence) are properly entertained as the strongest (and the weakest) possible connections between the members of these classes. Thus, the identification of "uncertainties" (in a similar sense) with probabilities between zero and one inclusively promises to provide an interpretation of the probabilistic hypothesis h_4 which might support a realistic interpretation of ontic indeterminism.

Because probabilities as frequencies are physical properties of physical systems, the frequency approach exerts considerable prima facie appeal as an appropriate conception within the quantum context. But appearances often deceive us, and this is no exception. For in order to apply to any particular quantum outcome, such as impacting in area B, it is essential to assign each quantum experiment, such as each emission of an electron from the source A, to the proper class of trials of some kind, where the kind of trial involved is determined by the complete set of relevant properties attending its occurrence, relative to the *frequency criterion of statistical relevance*. According to this criterion of relevance, any property F with respect to which frequencies differ, i.e., such that $P(B/A \cap F) \neq P(B/A \cap -F)$, is statistically relevant, necessarily. Unfortunately, there are excellent grounds for assuming that, for any two events, say, e_1 and e_2, there is at least one property, F, such that F is an aspect of e_1 but not of e_2. But this means that, even if e_1 and e_2 both happen to be emissions of electrons from the same source, A, nevertheless, the only circumstance under which they can both be classified as members of the same reference class, such as A, under the frequency criterion of statistical relevance, is when they both happen to have the same outcome, such as B. Hence, unless a quantum experiment, e_i, is assigned to a reference class in which its outcome, B, occurs with "certainty", it is not even logically possible that experiment e_i has been assigned to the proper class of trials.

In order to avoid misunderstanding, we must keep in mind that, whenever an outcome B, $C, \ldots,$ occurs, a question arises as to what set of properties A, $A \cap F^1$, $A \cap F^1 \cap F^2, \ldots,$ so to speak, "brought it about" (or, in other words, to what reference class the single event, e_i, should be assigned). For any difference in outcomes, of course, it is entirely possible that that difference was "brought about" by the occurrence of different relevant conditions. If any two events, e_i and e_j, differ with respect to at least one property F, however, then every event, strictly speaking, qualifies as an event of a distinctive *kind*, since, with respect to any other event, there must be at least one property, F^k, that is an aspect of e_i, while – F^k is an aspect of e_j. Intuitively, of course, we are inclined to believe that, even if two emissions of electrons from source A occurred at different times, under different weather conditions, in different colored rooms, and so forth, nevertheless, only some – but not all – of these conditions are explanatorily and/or predictively relevant with respect to a specific outcome, such as B, as the result of the specific event e_2, say, where e_2 belongs to many different reference classes, A, $A \cap F^1$, $A \cap F^1 \cap F^2, \ldots.$ With the frequency criterion of statistical relevance, however, there is no latitude for judgements that some of these properties are, but some of these properties are not, explanatorily/predictively relevant to a specific result, other than those displayed by frequencies *per se*: when "effects" differ, their "causes" differ, necessarily. Since every single event, e_i, is different in kind from every other, e_j, if a certain result, such as B, occurs on one trial, while an incompatible outcome, such as C, occurs on another, then the frequency criterion of statistical relevance, in principle, *demands* that the differences between these events have to qualify as statistically relevant. Indeed, the frequency interpretation thus understood cannot sustain probability assignments of other than zero and one. (For detailed elaboration, see Fetzer 1981, Chapter 4.)

Now if the only probabilities that are properly assignable to quantum outcomes are "certainties", i.e., degenerate probabilities of zero and of one, which actually represent constant non-occurrence and constant conjunction, respectively, then if this result follows as a matter of logical necessity from essential features of the frequency interpretation, then it is not logically possible for a probabilistic hypothesis – such as one assigning the probability of $\frac{1}{4}$ to landing in area B as an outcome of an emission from source A – to be *true*, when the probabilities involved are not degenerate probabilities of zero or of one. Moreover, since this conclusion is a necessary consequence of the frequency interpretation, it applies alike to events in the future as well as to events in the past, completely independently of our knowledge or belief or awareness thereof. But if this is the case, then it is logically impossible for the frequency interpretation to provide the theoretical foundation for a realistic interpretation of ontic indeterminism which the probabilistic hypothesis, h_4, requires. So let us assume that this approach is not what we desire.

According to the subjective conception, probabilities are supposed to be degrees of belief in particular propositions, such as, say, that an electron will impact in area B, when an individual x possesses certain other beliefs, such as that an electron has been emitted from source A. The ascription of a degree of belief thus assumes the following logical form,

(8) $[P_A(B) = r]_{xt}$

where 'r' denotes the degree of belief in the proposition that B when individual x already possesses the belief that A. The presence of subscripts, 'xt', of course, reflects that these probabilities are properties of individuals x at times t, where it is perfectly permissible for two different individuals, say, x and y, to have different degrees of belief in the same proposition, B, even when they are under the same conditions of belief, A, so long as their respective distributions of degrees of belief at any time t remain formally consistent (or "coherent"). Nevertheless, each individual x is required to preserve a special relationship between his distribution of degrees of belief at time t_i (before acquiring the belief that A) and at time t_j (after acquiring the belief that A), where his prior degree of belief in B conditional upon A, when x already holds other beliefs, say, F, is supposed to equal his posterior belief in B, when he has acquired the belief A as an addition to his other beliefs, F, i.e., $P_{F \cup A}(B) = P_F(B/A)$, a process known as "conditionalization" regulating degree of belief distributions across time. (Compare, for example, Fetzer 1981, Chapters 8 and 10.)

Since probabilities as degrees of belief are mental properties of particular persons-at-times, this interpretation does not appear to hold great promise as an appropriate conception within the quantum context. The probability with which an electron emitted from source A lands in area B, for example, need not be equal to $\frac{1}{4}$ for different individuals at the same time or even for the same individual at different times. The assignment of degrees of belief, of course, is determined by the complete set of relevant beliefs for an individual x at a time t, relative to a *subjective criterion of evidential relevance*, according to which any member of the set of beliefs that x accepts at t, such as A, is evidentially relevant to the degree of belief x assigns at t to some other belief, such as B, if the truth or falsity of A "makes a difference" to that degree of belief, i.e., $P_{F \cup A}(B) \neq P_{F \cup -A}(B)$. Now while this criterion of relevance permits the possibility that particular propositions, such as B, might be assigned degrees of belief which are not equal to degenerate probabilities of zero (as "incredulity") or of one (as "indubitability"), it does *not* permit the possibility that any outcome of which x becomes aware could be assigned any degree of belief other than equal to one. For x's prior degree of belief in B conditional upon B when x already holds other beliefs, F, has to be equal to one, as a function of coherence; but then x's posterior belief in B, when he has acquired the belief B as an addition to his other beliefs, F, also has to equal one, as a function of conditionalization, where $P_{F \cup B}(B) = P_F(B/B)$.

Not the least of the consequences attending this conception, therefore, is that, when probabilities are understood as degrees of belief, it is only possible for a quantum outcome to be assigned a probability value between zero and one so long as that specific outcome, such as B, remains unknown; for as soon as x becomes aware that B is the case, his degree of belief in the proposition that B has to change to one! This situation is illustrated by the Schrödinger "Cat Paradox", in fact, where the discovery that the cat is dead (or is alive) brings about a "reduction of the wave packet" in the form of an abrupt shift in subjective probabilities from $\frac{1}{2}$, let us say, to one! This phenomenon is not at all peculiar to quantum contexts, moreover, but commonplace with respect to tosses of coins, throws of dice, and "games of chance", in general, which are usually ontically deterministic – and only epistemically indeterministic as functions of ignorance. Although the subjective interpretation thus contributes toward clarification of the anomalous character of the "reduction of the wave packet", it does not appear to provide the theoretical foundation for a realistic interpretation of ontic indeterminism which the probabilistic hypothesis, h_4, requires. So let us assume that this approach too is not what we desire.

According to the propensity conception, finally, probabilities are to be understood as dispositional properties (or "causal tendencies") for one or another possible outcome, such as impacting in area B, in area C, etc., to be brought about (with a certain strength) by the occurrence of a test, trial, or experiment with a fixed kind of apparatus or arrangement, such as the emission of an electron from a source such as A. A probabilistic hypothesis under the propensity interpretation may be characterized as possessing the following logical form,

(9) $(x)(t)(Axt \ni_m Bxt^*)$

where 'm' denotes the strength of the tendency for an outcome of kind B at time t^* to be brought about by a single trial of kind A at time t – and t^* is equal to $t + \Delta t$. If the probability with which an electron x emitted from source A lands in area B is given by the square of the absolute value of the corresponding psi-function, $|\Psi_B|^2 = \frac{1}{4}$, as before, this means that a single trial of kind A would bring about an outcome of kind B with strength equal to $\frac{1}{4}$, which in turn (probabilistically) implies that a very large number of trials of kind A would tend to bring about an outcome of kind B with a relative frequency equal to $\frac{1}{4}$, where the strength of this tendency becomes enormously strong as the length of such a sequence increases without bound.

Since propensities, like frequencies, are properties of the world independently of anyone's knowledge, belief, or awareness thereof, we are again confronted by an intriguing point of view. In order to apply to any particular quantum outcome, such as impacting in area B, of course, it is essential to characterize each single trial, such as each emission of an electron from its source A, as a trial of the appropriate kind, where the kind involved is determined by the complete set of relevant properties attending its occurrence, as before, but now relative to the *propensity criterion of causal relevance*. According to this criterion of relevance, any property, F, with respect to which propensities differ, i.e., such that $(x)(t)[(Axt \cdot Fxt) \ni_m Bxt^*]$ and $(x)(t)[(Axt \cdot -Fxt) \ni_n Bxt^*]$, where $m \neq n$, is causally relevant to an outcome of kind B on a trial of kind A, necessarily. Since causally relevant properties need not be statistically relevant properties, the complete set of *relevant* properties present on a single trial can be a subset of the complete set of *properties* present at that trial, which means that the strength of the tendency for an outcome of kind B to be the effect of a trial of kind A can have a value between zero and one inclusively, where it is not the case that the only probabilities that are properly assignable to quantum outcomes are "certainties", i.e., degenerate probabilities with the values of zero and of one. (For further discussion, see Fetzer 1981.)

Like the frequency conception, the propensity conception is such that probabilities of zero and of one are logically compatible with any number of exceptions; consequently, "certainties" should be identified, not with constant non-occurrence and constant conjunction, but rather with null and universal strength instead. Thus, causal tendencies of universal strength obtain where, if only the relevant parameters are completely specified, it is possible, in principle, to predict the outcome that will occur in every single case without exception, assuming, of course, some such theory is true, because this situation is one of ontic determinism. And causal tendencies of probabilistic strength obtain where, if only the relevant parameters are completely specified, it is possible, in principle, not to predict the outcomes that occur in every single case without exception, but rather to derive probabilistic expectations for outcome distributions instead, assuming some such theory is true, because this situation is one of ontic indeterminism. Hence, the propensity interpretation overcomes the difficulties confronting the frequency interpretation, because it is logically possible for a probabilistic hypothesis – such as one assigning the probability of $\frac{1}{4}$ to landing in area B as an outcome of an emission from source A – to be *true*, even when the probabilities involved are not degenerate probabilities of zero or of one. And the propensity conception overcomes the difficulties confronting the subjective conception, because it is logically possible for a probabilistic hypothesis – such as one assigning the probability of $\frac{1}{4}$ to landing in area B as an outcome of an emission from source A – to be *true*, even when that outcome is known to have occurred. The situation thus appears to be as follows:

(10)

Three interpretations are possible:	Explanatory indeterminism	Predictive indeterminism
a. Frequency Constructions (FC):	No	No
b. Subjective Constructions (SC):	No	Yes
c. Propensity Constructions (PC):	Yes	Yes

Only the propensity interpretation permits probabilistic hypotheses assigning non-degenerate probabilities to known or unknown outcomes to be true. Since this is the case, moreover, it is logically possible for the propensity conception to provide the theoretical foundation for a realistic interpretation of ontic indeterminism which the probabilistic hypothesis, h_4, requires. So let us tentatively assume that this approach *is* what we desire.

Refining Einstein's Criterion

If these considerations are correct, then of the three interpretations which we have examined here – the frequency, the subjective, and the propensity – only the third accommodates the possibility, in principle, of providing a realistic interpretation of ontic indeterminism. Insofar as the arguments of the third section, which were based upon Einstein's criterion, suggested the conclusion that a realistic interpretation of ontic indeterminism might be a theoretical impossibility, such an interpretation must satisfy another criterion of reality than the one endorsed by Einstein for this reason:

(11) If, without in any way disturbing a system S, we can predict, not with certainty but with some probability less than 1, the value of a physical quantity, then either (a^*) our description of system S is not complete (in the sense of Assumption E) or (b^*) system S is not governed by strict causal laws (in the sense of Assumption D).

That (b^*) should be the case, of course, is entirely unproblematic, insofar as indeterministic situations are not governed by strict causal laws. But that (a^*) should be the case is highly problematic, since it invites the re-introduction of the hidden-variable hypothesis h_5. Fortunately, Feynman has supplied the missing element in the form of an alternative conception, which we shall refer to as "*Assumption F*" (for "Feynman"), as follows:

Assumption F: An ideal experiment is one in which all of the initial and final conditions of the experiment are completely specified. (Feynman et al. 1965, Vol. III, p. 1–10)

To be precise, an ideal experiment E is one for which all the *actual* initial conditions and *possible* final conditions are completely specified. I therefore suggest the following as an improvement upon Einstein's criterion:

(12) If an ideal experiment with a system S permits the prediction of its future states with deductive certainty or with probabilistic confidence, then that system is a deterministic or an indeterministic element of physical reality, respectively (in the sense of *Assumption F*)

where this criterion applies to indeterministic as well as to deterministic situations. And I further conclude that, while the Copenhagen Interpretation represents an unwarranted intrusion of subjectivism into physics, the propensity conception tends to remove (at least some of) the mystery from quantum phenomena, precisely as Popper has claimed.

References

Brush S (1979) Einstein and indeterminism. J Wash Acad Sci 69:89–94

Einstein A, Podolsky B, Rosen N (1935) Can quantum-mechanical description of reality be considered complete? Phys Rev, series 2, 47:777–780

Fetzer JH (1981) Scientific knowledge. D Reidel, Dordrecht

Feynman R, Leighton R, Sands M (1965) The Feynman lectures on physics, vol III. Addison-Wesley, Reading, MA

Folse HJ (1977) Complementarity and the description of experience. Int Philos Quart 17:377–392

Hooker CA (1972) The nature of quantum mechanical reality. In: Colodny RG (ed) Paradigms and paradoxes. University of Pittsburgh Press, Pittsburgh, pp 67–302
Popper KR (1957) The propensity interpretation of the calculus of probability, and the quantum theory. In: Korner S (ed) Observation and interpretation in the philosophy of physics. Dover, Delaware, pp 65–70
Popper KR (1967) Quantum mechanics without "The Observer". In: Bunge M (ed) Quantum theory and reality. Springer, New York, pp 7–44
Popper KR (1982) Quantum theory and the schism in physics. Rowman and Littlefield, Totowa, NJ
Reichenbach H (1944) Philosophic foundations of quantum mechanics. University of California Press, Berkeley and Los Angeles

Part IV
Probabilities in Inference and Decision

How Bayesian Confirmation Theory Handles the Paradox of the Ravens

Branden Fitelson and James Hawthorne

Introduction

The Paradox of the Ravens (aka, The Paradox of Confirmation) is indeed an old chestnut. A great many things have been written and said about this paradox and its implications for the logic of evidential support.[1] The first part of this paper will provide a brief survey of the early history of the paradox. This will include the original formulation of the paradox and the early responses of Hempel, Goodman, and Quine. The second part of the paper will describe attempts to resolve the paradox within a Bayesian framework, and show how to improve upon them. This part begins with a discussion of how probabilistic methods can help to clarify the statement of the paradox itself. And it describes some of the early responses to probabilistic explications. We then inspect the assumptions employed by traditional (canonical) Bayesian approaches to the paradox. These assumptions may appear to be overly strong. So, drawing on weaker assumptions, we formulate a new-and-improved Bayesian confirmation-theoretic resolution of the Paradox of the Ravens.

The Original Formulation of the Paradox

Traditionally, the Paradox of the Ravens is generated by the following two assumptions (or premises).

- **Nicod Condition (NC)**: For any object a and any predicate F and G, the proposition that a has both F and G confirms the proposition that every F has G. A more formal version of (NC) is the following claim: $(Fa{\cdot}Ga)$ confirms $(\forall x)(Fx \supset Gx)$, for any individual term 'a' and any pair of predicates 'F' and 'G'.

B. Fitelson (✉)
Dept. of Philosophy, University of California, CA 947202390, Berkeley
e-mail: branden@fitelson.org

J. Hawthorne
Dept. of Philosophy, University of Oklahoma, Norman, OK 73019, USA
e-mail: hawthorne@ou.edu

E. Eells and J.H. Fetzer (eds.), *The Place of Probability in Science*, Boston Studies in the Philosophy of Science 284, DOI 10.1007/978-90-481-3615-5_11,

- **Equivalence Condition (EC)**: For any propositions H_1, H_2, and E, if E confirms H_1 and H_1 is (classically) logically equivalent to H_2, then E confirms H_2.
  From (NC) and (EC), we can deduce the following, "paradoxical conclusion":
- **Paradoxical Conclusion (PC)**: The proposition that a is both non-black and a non-raven, $(\sim Ba \cdot \sim Ra)$, confirms the proposition that every raven is black, $(\forall x)(Rx \supset Bx)$.

The canonical derivation of (PC) from (EC) and (NC) proceeds as follows:

1. By (NC), $(\sim Ba \cdot \sim Ra)$ confirms $(\forall x)(\sim Bx \supset \sim Rx)$.
2. By Classical Logic, $(\forall x)(\sim Bx \supset \sim Rx)$ is equivalent to $(\forall x)(Rx \supset Bx)$.
3. By (1), (2), and (EC), $(\sim Ba \cdot \sim Ra)$ confirms $(\forall x)(Rx \supset Bx)$*QED*.

The earliest analyses of this infamous paradox were offered by Hempel, Goodman, and Quine. Let's take a look at how each of these famous philosophers attempted to resolve the paradox.

Early Analyses of the Paradox due to Hempel, Goodman, and Quine

The Analyses of Hempel and Goodman

Hempel (1945) and Goodman (1954) didn't view (PC) as paradoxical. Indeed, Hempel and Goodman viewed the argument above from (1) and (2) to (PC) as *sound*. So, as far as Hempel and Goodman are concerned, there is something misguided about whatever intuitions may have lead some philosophers to see "paradox" here. As Hempel explains (Goodman's discussion is very similar on this score), one might be misled into thinking that (PC) is false by conflating (PC) with a different claim (PC*) – a claim that is, in fact, false. Hempel warns us that [our emphasis]

> ... in the seemingly paradoxical cases of confirmation, we are often not judging the relation of the given evidence E *alone* to the hypothesis H ... instead, we tacitly introduce a comparison of H with a body of evidence which consists of E in conjunction with an additional amount of information we happen to have at our disposal.

We will postpone discussion of this crucial remark of Hempel's until the later sections on Bayesian clarifications of the paradox – where its meaning and significance will become clearer. Meanwhile, it is important to note that Hempel and Goodman also provide independent motivation for premise (1) of the canonical derivation of (PC) – a motivation independent of (NC) – in an attempt to further bolster the traditional argument.[2] The following argument for premise (1) is endorsed by both Hempel and Goodman [our emphasis and brackets]:

> If the evidence E consists *only* of one object which ... is a non-raven [$\sim Ra$], then E may reasonably be said to confirm that all objects are non-ravens [$(\forall x) \sim Rx$], and *a fortiori*, E supports the weaker assertion that all non-black objects are non-ravens [$(\forall x)(\sim Bx \supset \sim Rx)$].

This alternative argument for premise (1) presupposes the *Special Consequence Condition*:

(SCC) For all propositions H_1, H_2, and E, if E confirms H_1, and H_1 (classically) logically entails H_2, then E confirms H_2.

Early instantial and probabilistic theories of confirmation (e.g., those presupposed by Hempel et al. (1950)) embraced (SCC). But, from the point of view of contemporary Bayesian confirmation theory, (SCC) is false, as was first shown by Carnap (1950). We will return to this recent dialectic below, in our discussion of the paradox within the context of contemporary Bayesian confirmation theory. But before making the transition to Bayesian confirmation, let us briefly discuss Quine's rather influential response to the paradox, which deviates significantly from the views of Hempel and Goodman.

Quine on the Paradox of the Ravens

In his influential paper "Natural Kinds", Quine (1969) offers an analysis of the paradox of confirmation that deviates radically from the Hempel–Goodman line. Unlike Hempel and Goodman, Quine *rejects* the paradoxical conclusion (PC). Since Quine accepts classical logic, this forces him to reject either premise (1) or premise (2) of the (classically valid) canonical argument for (PC). Since Quine also accepts the (classical) equivalence condition (EC), he must accept premise (2). Thus, he is led, inevitably, to the rejection of premise (1). This means he must reject (NC) – and he does so. Indeed, according to Quine, not only does $(\sim Ba \cdot \sim Ra)$ *fail to confirm* $(\forall x)(\sim Bx \supset \sim Rx)$, but also $\sim Ra$ *fails to confirm* $(\forall x) \sim Rx$. According to Quine, the failure of instantial confirmation in these cases stems from the fact that the predicates 'non-black' $[\sim B]$ and 'non-raven' $[\sim R]$ are not *natural kinds* – i.e., the objects falling under $\sim B$ and $\sim R$ are not sufficiently similar to warrant instantial confirmation of universal laws involving $\sim B$ or $\sim R$. Only instances falling under natural kinds can warrant instantial confirmation of universal laws. Thus, for Quine, (NC) is the source of the problem here. He suggests that the *unrestricted* version (NC) is false, and must be replaced by a *restricted* version that applies only to natural kinds:

Quine–Nicod Condition (QNC): For any object a and any *natural kinds* F and G, the proposition that a has both F and G confirms the proposition that every F has G. More formally, $(Fa \cdot Ga)$ confirms $(\forall x)(Fx \supset Gx)$, for any individual term a, provided that the predicates 'F' and 'G' refer to *natural kinds*.

To summarize, Quine thinks (PC) is false, and that the (valid) canonical argument for (PC) is unsound because (NC) is false. Furthermore, according to Quine, once (NC) is restricted in scope to *natural kinds*, the resulting restricted instantial confirmation principle (QNC) is true, but useless for deducing (PC).[3] However, many other commentators have taken (NC) to be the real culprit here, as we'll soon see. We think that the real problems with (NC) (and (QNC)!) only become clear when the paradox is cast in more precise Bayesian terms, in a way that will be explicated in the second part of this paper. But we will first show how the Bayesian framework allows us to clarify the paradox and the historical debates surrounding it.

Bayesian Clarifications of (NC) and (PC)

Hempel (1945) provided a cautionary remark about the paradox. He warned us not to conflate the paradoxical conclusion (PC) with a distinct (intuitively) *false* conclusion (PC*) that (intuitively) does *not* follow from (NC) and (EC). We think Hempel's intuitive contrast between (PC) and (PC*) is important for a proper understanding the paradox. So, we'll discuss it briefly.

What, precisely, was the content of this (PC*)? Well, that turns out to be a bit difficult to say from the perspective of traditional, deductive accounts of confirmation. Based on the rest of Hempel's discussion and the penetrating recent exegesis of Patrick Maher (Maher 1999), we think the most accurate informal way to characterize (PC*) is as follows:

> (PC*) If one observes that an object a – *already known to be a non-raven* – is non-black (hence, is a non-black non-raven), then *this* observation confirms that all ravens are black.

As Maher points out, it is somewhat tempting to conflate (PC*) and (PC). But, Hempel did *not* believe that (PC*) was true (intuitively) about confirmation, nor did he think that (PC*) (intuitively) follows from (NC) and (EC). This is because, intuitively, *observing (known) non-ravens does not tell us anything about the color of ravens*. While this seems intuitively quite plausible, it is quite difficult to see how Hempel's *confirmation theory* can *theoretically ground* the desired distinction between (PC) and (PC*). What Hempel *says* is that we should not look at the evidence E in *conjunction* with other information that we might have at our disposal. Rather, we should look at the confirmational impact of learning *E and only E*.

There are two problems with this (the second worse than the first). First, as we have cast it (and as we think it *should* be cast), (PC*) is not a claim about the confirmational impact on $(\forall x)(Rx \supset Bx)$ of learning $\sim Ba$ in *conjunction* with other information about a (i.e., $\sim Ra$), but the impact on $(\forall x)(Rx \supset Bx)$ of learning $\sim Ba$ *given that you already know* $\sim Ra$. Basically, we are distinguishing the following two kinds of claims:

- E confirms H, *given* A – e.g., $\sim Ba$ confirms $(\forall x)(Rx \supset Bx)$, *given* $\sim Ra$ – *versus*
- $(E \cdot A)$ confirms H, *unconditionally* – e.g., $(\sim Ba \cdot \sim Ra)$ confirms $(\forall x)(Rx \supset Bx)$ *unconditionally*.

Note: in classical deductive logic, there *is no distinction* between:

- X entails Y, *given* Z, and
- $(X \cdot Z)$ entails Y

For this reason, Hempel's theory of confirmation (which is based on deductive entailment – see below) is incapable of making such a distinction. Perhaps this explains why he states things in terms of *conjunction*, rather than *conditionalization*. After all, he offers no confirmation-theoretical distinction between 'and' and 'given that'. So, while it seems that there is an *intuitive* distinction of the desired kind between (PC) and (PC*), it is unclear how Hempel's *theory* is supposed to make this distinction formally precise (see Maher (1999) for discussion).[4]

The second problem with Hempel's intuitive "explaining away" of the paradox is far more worrisome. As it turns out, Hempel's official theory of confirmation is *logically incompatible* with his intuitive characterization of what is going on. According to Hempel's theory of confirmation, the confirmation relation is *monotonic*. That is, Hempel's theory entails:

(M) If E confirms H, relative to *no* (or tautological) background information, then E confirms H relative to *any* collection of background information whatsoever.

The reason Hempel's theory entails (M) is that it explicates "E confirms H relative to K" as "E & K entails X", where the proposition X is obtained from the syntax of H and E in a certain complex way, which Hempel specifies (the technical details of Hempel's approach to confirmation won't matter for present purposes). Of course, if *E by itself* entails X, then so does E & K, for any K.[5] Thus, according to Hempel's theory of confirmation, if (PC) is true, then (PC*) must also be true. So, while intuitively compelling and explanatory, Hempel's suggestion that (PC) is true but (PC*) is false *contradicts* his own theory of confirmation. As far as we know, this logical inconsistency in Hempel (and Goodman's) discussions of the paradox of confirmation has not been discussed in the literature.[6]

It is clear that Hempel was onto something important here with his intuitive distinction between claims (PC) and (PC*), but his confirmation theory just lacked the resources to properly spell out his intuitions. Here contemporary Bayesian confirmation theory really comes in handy.

According to Bayesian confirmation theory, "E confirms H, given K", and "$(E \cdot K)$ confirms H, *unconditionally*" have quite different meanings. Essentially, this is possible because Bayesian explications of the confirmation relation do *not* entail *monotonicity* (M). Specifically, contemporary Bayesians offer the following account of conditional and unconditional confirmation – where hereafter, we will use the words "confirms" and "confirmation" in accordance with this Bayesian account:

- **Bayesian Confirmation** E confirms H, *given* K (or *relative to* K), just in case $\mathrm{P}[H|E \cdot K] > \mathrm{P}[H|K]$. And, E confirms H, *unconditionally*, just in case $\mathrm{P}[H|E] > \mathrm{P}[H]$, where $\mathrm{P}[\cdot]$ is some suitable probability function.[7]

It is easy to see, on this account of (conditional and unconditional) confirmation, that there will be a natural distinction between (PC) and (PC*). From a Bayesian point of view this distinction becomes:

(PC) $\mathrm{P}[(\forall x)(Rx \supset Bx)|{\sim}Ba \cdot {\sim}Ra] > \mathrm{P}[(\forall x)(Rx \supset Bx)]$, and
(PC*) $\mathrm{P}[(\forall x)(Rx \supset Bx)|{\sim}Ba \cdot {\sim}Ra] > \mathrm{P}[(\forall x)(Rx \supset Bx)|{\sim}Ra]$

What Hempel had in mind (charitably) is the former, *not* the latter. This is crucial for understanding the ensuing historical dialectic regarding the paradox. The important point here is that Bayesian confirmation theory has the theoretical resources to distinguish conditional and unconditional confirmation, but traditional (classical) deductive accounts do not. As a result Bayesian theory allows us to precisely articulate Hempel's intuition concerning why people might (falsely) believe that the paradoxical conclusion (PC) is false by conflating it with (PC*).

A key insight of Bayesian confirmation theory is that it represents confirmation as a *three*-place relation between *evidence E*, *hypothesis H*, and *background corpus K*. From this perspective the traditional formulation of the paradox is *imprecise* in an important respect: *it leaves unclear which background corpus is presupposed* in the (NC) – and, as a result, also in the (PC). In other words, there is a *missing quantifier* in the traditional formulations of (NC) and (PC). Here are four possible precisifications of (NC) (the corresponding precisifications of (PC) should be obvious):

- (NC_w) For any individual term 'a' and any pair of predicates 'F' and 'G', there is some possible background K such that $(Fa \cdot Ga)$ confirms $(\forall x)(Fx \supset Gx)$, *given K*.
- (NC_α) Relative to our *actual* background corpus K_α, for any individual term 'a' and any pair of predicates 'F' and 'G', $(Fa \cdot Ga)$ confirms $(\forall x)(Fx \supset Gx)$, *given* K_α.
- ($NC_\top$) Relative to *tautological* (or a priori) background corpus $K_\top$, for any individual term 'a' and any pair of predicates 'F' and 'G', $(Fa \cdot Ga)$ confirms $(\forall x)(Fx \supset Gx)$, *given* $K_\top$.
- (NC_s) Relative to *any possible* background corpus K, for any individual term 'a' and any pair of predicates 'F' and 'G', $(Fa \cdot Ga)$ confirms $(\forall x)(Fx \supset Gx)$, *given K*.

Which rendition of (NC) is the one Hempel and Goodman had in mind? Well, (NC_w) seems *too weak* to be of much use. There is bound to be *some* corpus with respect to which non-black non-ravens confirm 'All non-black things are non-ravens', but this corpus may not be very interesting (e.g., the corpus which contains '$(\sim Ba \cdot \sim Ra) \supset (\forall x)(\sim Bx \supset \sim Rx)$'!).

What about (NC_α)? Well, that depends. If we happen to (actually) *already know* that $\sim Ra$, then all bets are off as to whether $\sim Ba$ confirms $(\forall x)(\sim Bx \supset \sim Rx)$, relative to K_α (as Hempel suggests, and Maher makes precise). So, only a suitably restricted version of (NC_α) would satisfy Hempel's constraint. (We'll return to this issue, below.)

How about (NC_s)? This rendition is *too strong*. As we'll soon see, I.J. Good *demonstrated* that (NC_s) is *false* in a Bayesian framework.

What about ($NC_\top$)? As Maher (1999) skillfully explains, Hempel and Goodman (and Quine) have something much closer to ($NC_\top$) in mind. Originally, the question was whether learning *only* $(\sim Ba \cdot \sim Ra)$ *and nothing else* confirms that all ravens are black. And, it seems natural to understand this in terms of confirmation relative to "tautological (or a priori) background". We will return to the notion of "tautological confirmation", and the (NC_α) *vs* ($NC_\top$) controversy, below. But, first, it is useful to discuss I.J. Good's knock-down counterexample to (NC_s), and his later (rather lame) attempt to formulate a counterexample to ($NC_\top$).

I.J. Good's Counterexample to (NC_s) and His "Counterexample" to ($NC_\top$)

Good (1967) asks us to consider the following example (we're paraphrasing here):

- Our background corpus K says that exactly one of the following hypotheses is true: (H) there are 100 black ravens, no non-black ravens, and 1 million other birds, or else ($\sim H$) there are 1,000 black ravens, 1 white raven, and 1 million other birds. And K also states that an object a is selected at random from all the birds. Given this background K, we have:

$$\mathrm{P}[Ra \cdot Ba | (\forall x)(Rx \supset Bx) \cdot K] = \frac{100}{1000100}$$
$$< \mathrm{P}[Ra \cdot Ba | \sim(\forall x)(Rx \supset Bx) \cdot K] = \frac{1000}{1001000}$$

Hence, Good has described a background corpus K relative to which ($Ra \cdot Ba$) *dis*confirms $(\forall x)(Rx \supset Bx)$. This is sufficient to show that (NC_s) is false.

Hempel (1967) responded to Good by claiming that (NC_s) is *not* what he had in mind, since it smuggles too much "unkosher" (a posteriori) empirical knowledge into K. Hempel's challenge to Good was (again, charitably) to find a counterexample to ($NC_\top$). Good (1968) responded to Hempel's challenge with the following much less conclusive (rather *lame*, we think) "counterexample" to ($NC_\top$) (our brackets):

> ... imagine an infinitely intelligent newborn baby having built-in neural circuits enabling him to deal with formal logic, English syntax, and subjective probability. He might now argue, after defining a [raven] in detail, that it is initially extremely unlikely that there are any [ravens], and therefore that it is extremely likely that all [ravens] are black. ... On the other hand, if there are [ravens], then there is a reasonable chance that they are a variety of colours. Therefore, if I were to discover that even a black [raven] exists I would consider [$(\forall x)(Rx \supset Bx)$] to be less probable than it was initially.

Needless to say, this "counterexample" to ($NC_\top$) is far from conclusive! To us it seems completely unconvincing (see Maher (1999) for a trenchant analysis of this example). The problem here is that in order to give a rigorous and compelling counterexample to ($NC_\top$), one needs a *theory* of "tautological confirmation" – i.e. of "confirmation relative to tautological background". Good *doesn't have* such a theory (nor do most contemporary probabilists), which explains the lack of rigor and persuasiveness of "Good's Baby". However, Patrick Maher does have such an account; and he has applied it in his recent, neo-Carnapian, Bayesian analysis of the paradox of the ravens.

Maher's Neo-Carnapian Analysis of the Ravens Paradox

Carnap (1950, 1952, 1971, 1980) proposed various theories of "tautological confirmation" in terms of "logical probability". Recently Patrick Maher (1999, 2004) has brought a Carnapian approach to bear on the ravens paradox, with some very enlightening results. For our purposes it is useful to emphasize two consequences of Maher's neo-Carnapian, Bayesian analysis of the paradox. First, Maher shows that (PC^*) is *false* on a neo-Carnapian theory of (Bayesian) confirmation. That is, if we take a suitable class of Carnapian probability functions $P_c(\cdot|\cdot)$ – e.g., either those of Maher (1999) or Maher (2004) – as our "probabilities relative to tautological background", then we get the following result (see Maher 1999)

- $P_c[(\forall x)(Rx \supset Bx)|{\sim}Ba \cdot {\sim}Ra] = P_c[(\forall x)(Rx \supset Bx)|{\sim}Ra]$

Intuitively, this says that observing the color of (known) non-ravens tells us nothing about the color of ravens, relative to tautological background corpus. This is a theoretical vindication of Hempel's intuitive claim that (PC^*) is false – a vindication that is at best difficult to make out in Hempel's deductive theory of confirmation. But, all is not beer and skittles for Hempel.

More recently, Maher (2004) has convincingly argued (contrary to what he had previously argued in his (1999)) that, within a proper neo-Carnapian Bayesian framework, Hempel's ($NC_\top$) is false, and so is its Quinean "restriction" ($QNC_\top$). That is, Maher (2004) has shown that (from a Bayesian point of view) *pace* Hempel, Goodman, *and* Quine, *even relative to tautological background, positive instances do not necessarily confirm universal generalizations – not even for generalizations that involve only natural kinds*! The details of Maher's counterexample to ($QNC_\top$) (hence, to ($NC_\top$) as well) would take us too far afield. But, we mention it here because it shows that probabilistic approaches to confirmation are much richer and more powerful than traditional, deductive approaches. And, we think, Maher's work *finally* answers Hempel's challenge to Good – a challenge that went unanswered for nearly 40 years.

Moreover, Maher's results also suggest that Quine's analysis in "Natural Kinds" was off the mark. Contrary to what Quine suggests, the problem with (NC) is not merely that it needs to be restricted in scope to certain kinds of properties. The problems with (NC) run much deeper than that. Even the most promising Hempelian precisification of (NC) is false, and a restriction to "natural kinds" does not help (since Maher-style, neo-Carnapian counterexamples can be generated that employ only to "natural kinds" in Quine's sense).[8]

While Maher's neo-Carnapian analysis is very illuminating, it is by no means in the mainstream of contemporary Bayesian thought. Most contemporary Bayesians reject Carnapian logical probabilities and the Carnapian assumption that there is any such thing as "degree of confirmation relative to tautological background." Since contemporary Bayesians have largely rejected this project, they take a rather different tack to handle *the ravens* paradox.

The Canonical Contemporary Bayesian Approaches to the Paradox

Perhaps somewhat surprisingly, almost all contemporary Bayesians implicitly assume that the paradoxical conclusion is *true*. And, they aim only to "soften the impact" of (PC) by trying to establish certain *comparative* and/or *quantitative* confirmational claims. Specifically, Bayesians typically aim to show (at least) that the observation of a black raven, ($Ba \cdot Ra$), confirms "all ravens are black" *more strongly than* the observation of a non-black non-raven, ($\sim Ba \cdot \sim Ra$), relative to our *actual* background corpus K_α (which is assumed to contain no "unkosher" information about instance a). Specifically, most contemporary Bayesians aim to show (at least) that relative to some measure c of how strongly evidence supports a hypothesis, the following COMParative claim holds:[9]

$$(\text{COMP}_c)\quad c[(\forall x)(Rx \supset Bx), (Ra \cdot Ba)|K_\alpha] > c[(\forall x)(Rx \supset Bx),\ (\sim Ba \cdot \sim Ra)|K_\alpha].$$

Here $c(H, E|K)$ is some Bayesian measure of the *degree* to which E confirms H, relative to background corpus K. The typical Bayesian strategy is to isolate constraints on K_α that are as minimal as possible (hopefully, even ones that Hempel would see as kosher), but that guarantee that (COMP_c) obtains.

As it stands, (COMP_c) is somewhat unclear. There are *many* Bayesian relevance measures c that have been proposed and defended in the contemporary literature on Bayesian confirmation. The four most popular of these measures are the following (see Fitelson 1999, 2001 for historical surveys).[10]

- The Difference: $d[H, E|K] = \text{P}[H|E \cdot K] - \text{P}[H|K]$
- The Log-Ratio: $r[H, E|K] = \log(\text{P}[H|E \cdot K]/\text{P}[H|K])$
- The Log-Likelihood-Ratio: $l[H, E|K] = \log(\text{P}[E|H \cdot K]/\text{P}[E|\sim H \cdot K])$
- The Normalized Difference: $s[H, E|K] = \text{P}[H|E \cdot K] - \text{P}[H|\sim E \cdot K]$

Measures *d, r*, and l all satisfy the following *desideratum*, for all H, E_1, E_2, and K:

$$(\dagger)\ \text{if } \text{P}[H|E_1 \cdot K] > \text{P}[H|E_2 \cdot K],\ \text{then } c[H, E_1|K] > c[H, E_2|K].$$

But, interestingly, measure s does *not* satisfy (†). So, putting s aside, if one uses either *d, r*, or l to measure confirmation, then one can establish the desired comparative claim *simply* by demonstrating that:

$$(\text{COMP}_\text{P})\ \text{P}[(\forall x)(Rx \supset Bx)|Ra \cdot Ba \cdot K_\alpha] > \text{P}[(\forall x)(Rx \supset Bx)|\sim Ba \cdot \sim Ra \cdot K_\alpha]$$

(If one uses s, then one has a bit more work to do to establish the desired comparative conclusion, because (COMP_P) does not entail (COMP_s).)[11]

Some Bayesians go farther than this by trying to establish not only the *comparative* claim (COMPc), but also the *quantitative* claim that the observation of a non-black non-raven confirms "All ravens are black" to a very minute degree. That is, in addition to the comparative claim, some Bayesians also go for the following QUANTative claim:

$$(\mathrm{QUANT}_c) \quad c[(\forall x)(Rx \supset Bx), ({\sim}Ba \cdot {\sim}Ra)|K_\alpha] > 0, \text{ but very nearly } 0.$$

Let's begin by discussing the canonical contemporary Bayesian *comparative* analysis of the paradox. In essence, almost all such accounts trade on the following three assumptions about K_α (where we may suppose that the object a is sampled at random from the universe):[12]

(1) $\mathrm{P}[{\sim}Ba|K_\alpha] > \mathrm{P}[Ra|K_\alpha]$.
(2) $\mathrm{P}[Ra|(\forall x)(Rx \supset Bx) \cdot K_\alpha] = \mathrm{P}[Ra|K_\alpha]$.
(3) $\mathrm{P}[{\sim}Ba|(\forall x)(Rx \supset Bx) \cdot K_\alpha] = \mathrm{P}[{\sim}Ba|K_\alpha]$.

Basically, assumption (1) relies on our knowledge that (according to K_α) there are more non-black objects in the universe than there are ravens. This seems like a very plausible distributional constraint on K_α, since – as far as we *actually know* – it is true. Assumptions (2) and (3) are more controversial. We will say more about them shortly. First, we note an important and pretty well-known theorem.

Theorem (1)–(3) entails ($\mathrm{COMP_P}$). Therefore, since *d, r*, and *l* each satisfy (†), it follows that (1)–(3) entails (COMP_d), ($\mathrm{COMP_r}$), and (COMP_l).

In fact, (1)–(3) entails *much more* than ($\mathrm{COMP_P}$), as the following theorem illustrates:

Theorem (1)–(3) also entail the following:

(4) $\mathrm{P}[(\forall x)(Rx \supset Bx)|{\sim}Ba \cdot {\sim}Ra \cdot K_\alpha] > \mathrm{P}[(\forall x)(Rx \supset Bx)|K_\alpha]$.
(5) $s[(\forall x)(Rx \supset Bx), (Ra \cdot Ba)|K_\alpha] > s[(\forall x)(Rx \supset Bx), ({\sim}Ba \cdot {\sim}Ra)|K_\alpha]$.

In other words, (4) tells us that assumptions (1)–(3) entail that the observation of a non-black non-raven *positively confirms* that all ravens are black – i.e., that the paradoxical conclusion (PC) is true. And, (5) tells us that even according to measure *s* (a measure that *violates* (†)) the observation of a black raven confirms that all ravens are black more strongly than the observation of a non-black non-raven.

The fact that (1)–(3) entail (4) and (5) indicates that the canonical Bayesian assumptions go far beyond the minimal comparative claim most Bayesians were looking for. Why, for instance, should a Bayesian be committed to the *qualitative* paradoxical conclusion (PC)? After all, as Patrick Maher and I.J. Good have made so clear, probabilists don't *have* to be committed to qualitative claims like (NC) and (PC). It would be nice (and perhaps more informative about the workings of Bayesian confirmation) if there were assumptions weaker than (1)–(3) that sufficed to establish (*just*) the *comparative* claim ($\mathrm{COMP_P}$), while implying no commitment to specific *qualitative* claims like (PC). Happily, there are such weaker conditions. But, before we turn to them, we first need to briefly discuss the *quantitative* Bayesian approaches as well.

Various Bayesians go farther than (COMP_c) in their analysis of the ravens paradox. They seek to identify stronger constraints, stronger background knowledge K_α, that entails *both* (COMP_c) *and* (QUANT_c). The most common strategy along these lines is simply to strengthen assumption (1), as follows:

(1′) $P[{\sim}Ba|K_\alpha] \gg P[Ra|K_\alpha]$ – e.g., because there are *far* fewer ravens than non-black things in the universe.

Peter Vranas (2004) provides a very detailed discussion of quantitative Bayesian approaches to the ravens paradox along these lines. We won't dwell too much on the details of these approaches here. Vranas has already done an excellent job of analyzing them. However, some brief remarks on a result Vranas proves (and uses in his analysis) are worth considering.

Vranas shows that assumptions (1′) and (3) (without (2)) are *sufficient* for (QUANT_c) to hold – i.e. for $(\forall x)(Rx \supset Bx)$ to be positively confirmed by $({\sim}Ba \cdot {\sim}Ra)$ given K_α, but only by a very small amount. He shows this for all four measures of confirmation *d, r, l*, and *s*. Moreover, he argues that in the presence of (1′), (3) is "approximately necessary" for (QUANT_c). That is, he *proves* that given (1′), and supposing that $P[H|K_\alpha]$ is not too small, the following approximate claim is *necessary* for (QUANT_c):

(3′) $P[{\sim}Ba|(\forall x)(Rx \supset Bx) \cdot K_\alpha] \approx P[{\sim}Ba|K_\alpha]$.

Vranas then argues that Bayesians have given no good reason for *assuming* this (necessary and sufficient) condition. Thus, he concludes, Bayesian resolutions of the paradox that claim non-black non-ravens confirm by a tiny bit, due to assumption (1′), have failed to establish a condition they must employ to establish this claim – they have failed to establish (3′).[13]

Vranas' claim that (3) is "approximately necessary" for (QUANT_c) may be somewhat misleading. It makes it sound as if (3) has a certain property. But, in fact, nothing about (3) itself follows from Vranas' results. It is more accurate to say (as Bob Dylan might) that "approximately (3)" (i.e., (3′)) is *necessary* for (QUANT_c). To see the point, note that (3) is a rather strong *independence* assumption, which entails many other identities, including:

(3.1) $P[(\forall x)(Rx \supset Bx)|Ba \cdot K_\alpha] = P[(\forall x)(Rx \supset Bx)|K_\alpha]$, and
(3.2) $P[(\forall x)(Rx \supset Bx)|Ba \cdot K_\alpha] = P[(\forall x)(Rx \supset Bx)|{\sim}Ba \cdot K_\alpha]$.

But, (3′) is *not* an independence assumption. Indeed, (3′) is far weaker than an independence assumption, and it does *not* entail the parallel *approximates*:

(3′.1) $P[(\forall x)(Rx \supset Bx)|Ba \cdot K_\alpha] \approx P[(\forall x)(Rx \supset Bx)|K_\alpha]$, and
(3′.2) $P[(\forall x)(Rx \supset Bx)|Ba \cdot K_\alpha] \approx P[(\forall x)(Rx \supset Bx)|{\sim}Ba \cdot K_\alpha]$.

Vranas argues convincingly that strong independence assumptions like (3) (and (2)) have not been well motivated by Bayesians who endorse the quantitative approach to the ravens paradox. He rightly claims that this is a lacuna in the canonical quantitative Bayesian analyses of the paradox. But, what he ultimately shows is somewhat weaker than appearances might suggest. In the next two sections we will describe (*pace* Vranas and most other commentators) considerably weaker sets of assumptions for the comparative and the quantitative Bayesian approaches.

A New Bayesian Approach to the Paradox

As we have seen, Bayesians typically make two quite strong independence assumptions in order to establish the comparative claim that a black raven confirms more than does a non-black non-raven. In addition they usually suppose that given only actual background knowledge K_α, a non-black instance is more probable than a raven instance. Happily, there is a quite satisfactory analysis of *the ravens* that employs none of these assumptions up front. This solution to the ravens paradox is more general than any other solution we know of, and it draws on much weaker assumptions. It *solves* the paradox in that it supplies plausible necessary and sufficient conditions for an instance of a black raven to be *more favorable* to 'All ravens are black' than an instance of a non-black non-raven. Our most general result doesn't depend on whether the Nicod Condition (NC) is satisfied, and does not draw on probabilistic independence. Nor does it assume that more plausible claim that (given background knowledge) a non-black instance is more probable than a raven instance (i.e. assumption (1) in the previous section). Indeed, the conditions for this result may be satisfied even if an instance of a black raven *lowers* the degree of confirmation for 'All ravens are black'. In that case it just shows that non-black non-ravens lower the degree of confirmation even more. Thus, this result strips the Bayesian solution to bare bones, decoupling it from any of the usual assumptions, and then permits the introduction of whatever additional suppositions may seem plausible and fitting (e.g. those leading to positive confirmation) to be added separately.

For the sake of notational simplicity, let 'H' abbreviate 'All ravens are black' – i.e., '$(\forall x)(Rx \supset Bx)$'. Let '$K$' be a statement of whatever background knowledge you may think relevant – e.g. K might imply, among other things, that ravens exist and that non-black things exist, $((\exists x)Rx \cdot (\exists x){\sim}Bx)$. One object, call it '$a$' will be observed for color and to see whether it is a raven. The idea is to assess, in advance of observing it, whether a's turning out to be a black raven, $(Ra \cdot Ba)$, would make H more strongly supported than would a's turning out to be a non-black non-raven, $({\sim}Ra \cdot {\sim}Ba)$. We want to find plausible conditions for $\mathrm{P}[H|Ba \cdot Ra \cdot K] > \mathrm{P}[H|{\sim}Ba \cdot {\sim}Ra \cdot K]$ to hold. Equivalently, we want to find plausible conditions for the ratio $\mathrm{P}[Ba \cdot Ra|H \cdot K]/\mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K]$ to exceed the ratio $\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|H \cdot K]/\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K]$.[14] We will attack the paradox by finding plausible sufficient and necessary conditions for this relationship between likelihood-ratios.[15] Notice that in general this relationship, $\mathrm{P}[Ba \cdot Ra|H \cdot K]/\mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K] > \mathrm{P}[{\sim}Ba \cdot {\sim}Ra|H \cdot \mathrm{K}]/\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K]$, may hold regardless of whether the instance $(Ba \cdot Ra)$ raises the confirmation of the hypothesis – i.e., regardless of whether $\mathrm{P}[H|Ba \cdot Ra \cdot K]$ is greater than, or less than, $P[H|K]$.[16] Thus, no condition that implies black ravens *raise* the degree of confirmation can be a *necessary condition* for black ravens to yield greater support than non-black non-ravens. Any such *positive confirmation implying condition* goes beyond what is strictly needed here.

We assume throughout the remainder of the paper the following very weak and highly plausible *non-triviality conditions*:

Non-triviality Assumptions: $P[Ba \cdot Ra|K] > 0$, $P[{\sim}Ba \cdot {\sim}Ra|K] > 0$, $P[{\sim}Ba \cdot Ra|K] > 0$, $0 < P[H|Ba \cdot Ra \cdot K] < 1$, $0 < P[H|{\sim}Ba \cdot {\sim}Ra \cdot K] < 1$.[17]

That is, we assume that it is at least epistemically (confirmationally) possible, given only background K, that observed object a will turn out to be a black raven; and possible that a will turn out to be a non-black non-raven; and even possible that a will turn out to be a non-black raven – a falsifying instance of H. Furthermore, we assume that finding a to be a black raven neither absolutely *proves* nor absolutely *falsifies* 'All ravens are black', nor does finding a to be a non-black non-raven do so.

Our analysis of *the ravens* will draw on three factors, which we label 'p', 'q', and 'r'.

Definition: *Define* $\mathrm{q} = P[{\sim}Ba|{\sim}H \cdot K]/P[Ra|{\sim}H \cdot K]$, *define* $\mathrm{r} = P[{\sim}Ba|H \cdot K]/P[Ra|H \cdot K]$, *and define* $\mathrm{p} = P[Ba|Ra \cdot {\sim}H \cdot K]$.

Given *Non-triviality*, p, q, and r are well-defined (q and r have non-zero denominators); q and r are greater than 0; and p is greater than 0 and less than 1. (See Lemma 1 in the Appendix.)

The factor r represents how much more likely it is that a will be a non-black thing than be a raven *if* the world in fact contains only black ravens (i.e. if H is true). Given the kind of world we think we live in, r should be quite large, since even if all of the ravens are black, the non-black things far outnumber the ravens. Similarly, the factor q represents how much more likely it is that a will be a non-black thing than be a raven *if* the world in fact contains non-black ravens (i.e. if H is false). Given the kind of world we think we live in, q should also be quite large, since the non-black things far outnumber the ravens even if some of the non-black things happen to be ravens. However, though plausibly r and q are very large, for now we will assume neither this nor anything else about their values except what is implied by the Non-triviality Assumptions – i.e. that r and q are well-defined and greater than 0.

Suppose that H is in fact false – i.e. non-black ravens exist – and suppose that a is a raven. How likely is it that a will turn out to be black? The factor p represents this likelihood. This factor may be thought of as effectively representing a "mixture" of the likelihoods due to the various possible alternative hypotheses about the frequency of black birds among the ravens. It would be reasonable to suppose that the value of p is pretty close to 1 – if there are non-black ravens, their proportion among all ravens is most plausibly some small percentage; so the proportion of black birds among ravens should be a fairly small increment below 1. However, for now we will not assume this, or anything else about the value of p, except what is implied by the Non-triviality Assumptions – i.e. that $0 < \mathrm{p} < 1$ (shown in Lemma 1 of the Appendix).

It turns out that the relative confirmational support for H from to a black raven instance as compared to that from a non-black non-raven instance is merely a function of p, q, and r.

Theorem 1. *Given Non-triviality, it follows that* $\mathrm{q} > (1-\mathrm{p}) > 0$ *and*

$$\frac{\mathrm{P}[Ba \cdot Ra|H \cdot K]\,/\mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K]}{\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|H \cdot K]\,/\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K]} = [\mathrm{q}-(1-\mathrm{p})]\,/\,(\mathrm{p}\cdot\mathrm{r}) > 0.$$

(This and the other theorems are proved in the Appendix.)

This theorem does not itself express the necessary and sufficient conditions for black ravens to favor ‘All ravens are black’ more strongly than do non-black non-ravens. But an obvious Corollary does so.

Corollary 1. *Given Non-triviality,*

$$\frac{\mathrm{P}[Ba \cdot Ra|H \cdot K]\,/\mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K]}{\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|H \cdot K]\,/\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K]} > 1 \text{ if and only if } \mathrm{q}-(1-\mathrm{p}) > \mathrm{p}\cdot\mathrm{r}.$$

And, more generally, for any real number s,

$$\frac{\mathrm{P}[Ba \cdot Ra|H \cdot K]\,/\mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K]}{\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|H \cdot K]\,/\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K]} = \mathrm{s} = [\mathrm{q}-(1-\mathrm{p})]\,/\,(\mathrm{p}\cdot\mathrm{r}) > 1 \text{ if and only if } [\mathrm{q}-(1-\mathrm{p})] = \mathrm{s}\cdot\mathrm{p}\cdot\mathrm{r} > \mathrm{p}\cdot\mathrm{r}.$$

This gives us a fairly useful handle on what it takes for a black raven to support H more than a non-black non-raven. For instance, suppose that $\mathrm{q} = \mathrm{r}$. Then the corollary implies that the value of the ratio of likelihood-ratios is greater than 1 *just in case* $\mathrm{q} = \mathrm{r} > 1$.[18] Thus, if the likelihood that an object is non-black is greater than the likelihood that it's a raven, and is greater by the same amount regardless of whether or not every raven is black, then a black raven supports ‘All ravens are black’ more strongly than does a non-black non-raven.

Notice that none of this depends on either *Ba* or *Ra* being probabilistically independent of H. Such independence, if it held, would make $\mathrm{P}[{\sim}Ba|{\sim}H \cdot K] = \mathrm{P}[{\sim}Ba|H \cdot K] = \mathrm{P}[{\sim}Ba|K]$ and make $\mathrm{P}[Ra|{\sim}H \cdot K] = \mathrm{P}[Ra|H \cdot K] = \mathrm{P}[Ra|K]$. In that case we would indeed have $\mathrm{q} = \mathrm{r}$, and so the result discussed in the previous paragraph would apply. However, that result applies even in cases where probabilistic independence fails miserably – even when $\mathrm{P}[{\sim}Ba|{\sim}H \cdot K]/\mathrm{P}[{\sim}Ba|H \cdot K]$ is very far from 1, provided only that $\mathrm{P}[Ra|{\sim}H \cdot K]/\mathrm{P}[Ra|H \cdot K]$ is equally far from 1.

What if $\mathrm{q} \neq \mathrm{r}$? Theorem 1 tell us that $\mathrm{q} > (1-\mathrm{p}) > 0$, so $\mathrm{q}-(1-\mathrm{p})$ is positive and a little smaller than q itself. As long as this $\mathrm{q}-(1-\mathrm{p})$ remains larger than r, the corollary tells us that the likelihood-ratio due to a black raven favors H more than does the likelihood-ratio due to a non-black non-raven. Indeed $\mathrm{q}-(1-\mathrm{p})$ need only remain larger than a fraction p of r in order to yield the desired result.

It turns out that 1/p is a convenient benchmark for comparing the size of the black-raven likelihood-ratio to the size non-black-non-raven likelihood-ratio.

Corollary 2. *Given Non-triviality, for real number s such that*

$$\frac{\mathrm{P}[Ba \cdot Ra|H \cdot K]/\mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K]}{\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|H \cdot K]/\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K]} = \mathrm{s} = [\mathrm{q} - (1 - \mathrm{p})]/(\mathrm{p} \cdot \mathrm{r}),$$

we have the following:

(1) $\mathrm{s} > (1/\mathrm{p}) > 1$ iff $\mathrm{q} - (1 - \mathrm{p}) > \mathrm{r}$.
(2) $\mathrm{s} = (1/\mathrm{p}) > 1$ iff $\mathrm{q} - (1 - \mathrm{p}) = \mathrm{r}$.
(3) $(1/\mathrm{p}) > \mathrm{s} > 1$ iff $\mathrm{r} > \mathrm{q} - (1 - \mathrm{p}) > \mathrm{p} \cdot \mathrm{r}$.

Notice that when $\mathrm{q} = \mathrm{r}$, Clause 3 applies (because then $\mathrm{r} > \mathrm{q} - (1 - \mathrm{p})$); so the value of the ratio of the likelihood-ratios, s, must be strictly between 1/p and 1. Alternatively, when q diminished by $(1 - \mathrm{p})$ is greater than r, Clause 1 applies; so the ratio of likelihood-ratios s must be greater than (1/p), possibly much greater. Indeed, looking back at Corollary 1, we see that the value of the ratio of likelihood ratios s can be enormous, provided only that $[\mathrm{q} - (1 - \mathrm{p})] \ggg (\mathrm{p} \cdot \mathrm{r})$.

The emergence of 1/p as a particularly useful benchmark is no accident. For, p is just $\mathrm{P}[Ba|Ra \cdot {\sim}H \cdot K]$, so $1/\mathrm{p} = \mathrm{P}[Ba|Ra \cdot H \cdot K]/\mathrm{P}[Ba|Ra \cdot {\sim}H \cdot K]$. Furthermore, if the usual independence assumption (2) were to hold (i.e. if $\mathrm{P}[Ra|H \cdot K] = \mathrm{P}[Ra|K]$), it would follow that $\mathrm{P}[Ra|H \cdot K] = \mathrm{P}[Ra|{\sim}H \cdot K]$; and then we'd have $1/\mathrm{p} = \mathrm{P}[Ba \cdot Ra|H \cdot K]/\mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K]$. Following this thought further, the usual Bayesian analysis adds independence assumption (3) (i.e. $\mathrm{P}[{\sim}Ba|H \cdot K] = \mathrm{P}[{\sim}Ba|K]$) to get $\mathrm{P}[{\sim}Ba|H \cdot K] = \mathrm{P}[{\sim}Ba|{\sim}H \cdot K]$; from which we'd have $\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|H \cdot K]/\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K] = \mathrm{P}[{\sim}Ba|H \cdot K]/(\mathrm{P}[{\sim}Ra|{\sim}Ba \cdot {\sim}H \cdot K] \cdot \mathrm{P}[{\sim}Ba|{\sim}H \cdot K]) = 1/\mathrm{P}[{\sim}Ra|{\sim}Ba \cdot {\sim}H \cdot K]$, where $\mathrm{P}[Ra|{\sim}Ba \cdot {\sim}H \cdot K]$ should be just a smidgen, ε, above 0 – because, very probably, only a really minuscule proportion of the non-black things are ravens, regardless of whether H is true or false. Thus, the usual analysis would peg the *ratio of likelihood-ratios* at a value $\mathrm{s} = (1 - \varepsilon)/\mathrm{p}$ (for ε almost 0), which is just a tiny bit below 1/p – which is only within the range of possible values for s encompassed by Clause 3 of Corollary 2, and merely within the uppermost end of that range. In light of this, the benchmark 1/p in Corollary 2 provides a telling indicator of the extent to which our treatment supersedes the usual approach.

Theorem 1 and its Corollaries show that for a very wide range of probabilistic confirmation functions P, a black raven is *more confirming* of 'All ravens are black' than is a non-black non-raven. These functions are so diverse that some of them even permit a black raven to provide evidence against 'All ravens are black' (i.e. make $\mathrm{P}[Ba \cdot Ra|H \cdot K]/\mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K] < 1$). Only a small range of these functions abide by the usual independence claims. For black ravens to be *more confirming*, all that matters are the relative sizes of q and r, as mediated by the factor p.

Let's now look at one more theorem that *solves* the paradox by drawing on additional conditions that restrict the values of q and r in a plausible way. This result is less general than Theorem 1 and its corollaries, but closely related to them.[19]

Theorem 2. *Given Non-triviality, both of the following clauses hold:*

(2.1) *If* $\mathrm{P}[{\sim}Ba|H \cdot K] > \mathrm{P}[Ra|H \cdot K]$ *(i.e. if* $\mathrm{r} > 1$*)* and $\mathrm{O}[H|Ra \cdot K]/\mathrm{O}[H|{\sim}Ba \cdot K] > \mathrm{p} + (1-\mathrm{p})/\mathrm{r}$ (where, 'O' is the *odds*), then

$$\frac{\mathrm{P}\left[Ba \cdot Ra|H \cdot K\right]/\mathrm{P}\left[Ba \cdot Ra|{\sim}H \cdot K\right]}{\mathrm{P}\left[{\sim}Ba \cdot {\sim}Ra|H \cdot K\right]/\mathrm{P}\left[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K\right]} > 1.$$

(2.2) *If* $\mathrm{P}[{\sim}Ba|H \cdot K] \leqslant \mathrm{P}[Ra|H \cdot K]$ *(i.e.* $\mathrm{r} \leqslant 1$*), but either* $\mathrm{P}[{\sim}Ba|K] > \mathrm{P}[Ra|K]$ *or (at least)* $\mathrm{P}[{\sim}Ba|{\sim}H \cdot K] > \mathrm{P}[Ra|{\sim}H \cdot K]$ *(i.e.* $\mathrm{q} > 1$*),* then

$$\frac{\mathrm{P}\left[Ba \cdot Ra|H \cdot K\right]/\mathrm{P}\left[Ba \cdot Ra|{\sim}H \cdot K\right]}{\mathrm{P}\left[{\sim}Ba \cdot {\sim}Ra|H \cdot K\right]/\mathrm{P}\left[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K\right]} > 1.$$

Clause (2.1) is the more interesting case, since its antecedent conditions are a better fit to the way we typically judge our world to be. The first antecedent of (2.1) draws on the idea that, provided all of the ravens are black, a randomly selected object a is more likely (in our world) to be a non-black thing than a raven. This seems really quite plausible. Indeed, not only does it seem that $\mathrm{P}[{\sim}Ba|H \cdot K]$ is *merely greater than* $\mathrm{P}[Ra|H \cdot K]$, quite plausibly $\mathrm{P}[Ra|H \cdot K]$ is close enough to 0 that $\mathrm{P}[{\sim}Ba|H \cdot K]$ is billions of times greater than $\mathrm{P}[Ra|H \cdot K]$ (though the theorem itself doesn't suppose *that*).

Now consider the second antecedent to (2.1). One wouldn't normally think that the mere fact that an object is black (without also taking account of whether it's a raven) should provide *more evidence* for 'All ravens are black' than would the mere fact that an object is a raven (without taking account of its color). Indeed, generally speaking, one would expect $\mathrm{O}[H|Ra \cdot K]$ to be very nearly equal to $\mathrm{O}[H|{\sim}Ba \cdot K]$. However, the second condition for Clause (2.1) is even weaker than this. Notice that for $\mathrm{r} > 1$ the term $\mathrm{p} + (1-\mathrm{p})/\mathrm{r}$ is less than $\mathrm{p} + (1-\mathrm{p}) = 1$; and the larger r happens to be (i.e. the greater the ratio $\mathrm{r} = \mathrm{P}[{\sim}Ba|H \cdot K]/\mathrm{P}[Ra|H \cdot K]$ is), the smaller $\mathrm{p} + (1-\mathrm{p})/\mathrm{r}$ will be, approaching the lower bound $\mathrm{p} = \mathrm{P}[Ba|Ra \cdot {\sim}H \cdot K]$ for very large r. Thus, the second condition for (2.1) will be satisfied provided that *either* $\mathrm{O}[H|Ra \cdot K]$ is bigger than or equal to $\mathrm{O}[H|{\sim}Ba \cdot K]$ (perhaps much bigger) *or* $\mathrm{O}[H|Ra \cdot K]$ is a bit smaller than $\mathrm{O}[H|{\sim}Ba \cdot K]$. Thus, this second condition can *fail to hold* only if (without taking account of whether it's a raven) a black object provides *more than a bit more evidence* for 'All ravens are black' than would a raven (without taking account of its color).[20]

Although the antecedent conditions for Clause (2.2) seem a less plausible fit to our world, it fills out Theorem 2 in an interesting way. Think of it like this. It is reasonable to suppose, given plausible background knowledge, that the non-black things will be much more numerous than ravens, regardless of whether all the ravens are black. But perhaps this intuition is confused. It is clearly guided by the fact that we inhabit a world in which there are far more non-black things than ravens. Problem is, if our world is one in which there are non-black ravens, we may only be warranted in taking the non-black things to outnumber the ravens in worlds like our – i.e. worlds where H is false. If, on the other hand, ours happens to be a world in which all of the ravens are black, then we may only be warranted in taking the non-black things to outnumber the ravens in worlds like ours – i.e. worlds where H is true. But we don't know which of these two kinds of worlds ours happens to be. That is precisely what is at issue – precisely what the evidence is supposed to tell us. Nevertheless, we can easily fineness this apparent difficulty. For, the apparent dilemma takes it as granted that *either* non-black things are much more numerous than ravens if H holds, *or* non-black things are much more numerous than ravens if $\sim H$ holds. Thus, given reasonable background knowledge, for an object a about which nothing else is known, *either* $\mathrm{P}[\sim Ba|H \cdot K] > \mathrm{P}[Ra|H \cdot K]$ (i.e. $\mathrm{r} > 1$) *or* $\mathrm{P}[\sim Ba|\sim H \cdot K] > \mathrm{P}[Ra|\sim H \cdot K]$ (i.e. $\mathrm{q} > 1$) (or perhaps $\mathrm{P}[\sim Ba|K] > \mathrm{P}[Ra|K]$). But Clause (2.1) of the theorem already takes account of the case where $\mathrm{P}[\sim Ba|H \cdot K] > \mathrm{P}[Ra|H \cdot K]$. So Clause (2.2) deals with the remaining case: that in case $\mathrm{P}[\sim Ba|H \cdot K] \leqslant \mathrm{P}[Ra|H \cdot K]$ (i.e. $\mathrm{r} \leqslant 1$) holds, at least $\mathrm{P}[\sim Ba|\sim H \cdot K] > \mathrm{P}[Ra|\sim H \cdot K]$, or maybe $\mathrm{P}[\sim Ba|K] > \mathrm{P}[Ra|K]$ holds. This is the only condition required for (2.2), and it's a very weak condition indeed.

Consider the disjunction of the antecedent conditions for Clause (2.1) with the antecedent conditions for Clause (2.2). This disjunction is a highly plausible claim – even more plausible than each antecedent taken alone. Given realistic background knowledge K, any reasonable probabilistic confirmation function P should surely satisfy the full antecedent of at least one of these two clauses. Thus, a black raven should *favor* 'All ravens are black' *more* than a non-black non-raven over a *very wide range* of circumstances. Furthermore, neither of the usual *approximate independence conditions* is required for this result. Thus, Theorem 1 and its corollaries together with Theorem 2 dissolve any air of a *qualitative paradox* in the case of *the ravens*.

Quantitative Results

Traditional *quantitative* Bayesian approaches also make rather strong independence-like assumptions. For example, in order to establish that a non-black non-raven positively confirms 'All ravens are black' by (only) a very small amount – the thesis we've labeled (QUANT_c), $c[H, \sim Ba \cdot \sim Ra|K] > 0$ but very near 0 – the usual approach employs an (at least approximate) independence assumption like (3) or (3′), $\mathrm{P}[\sim Ba|H \cdot K] \approx \mathrm{P}[\sim Ba|K]$, together with an assumption like (1′), $\mathrm{P}[\sim Ba|K_\alpha] \gg \mathrm{P}[Ra|K]$.[21]

Quantitative claims like (QUANT_c) are most informative when cashed out in terms of a specific measure of confirmation c. That is, although several of the well-studied measures of incremental confirmation (d, r, and l) agree with regard to *qualitative* confirmational relationships, their *quantitative* scales differ in ways that

that make quantitative results difficult to compare meaningfully across measures. So in this section we'll restrict our discussion to a single measure of incremental confirmation. In our judgment the most suitable Bayesian measure of incremental confirmation is the (log) likelihood-ratio measure.[22] We have detailed reasons for this assessment (see Fitelson 2001, 2004), but we'll not pause to discuss them here. Let's see what the likelihood-ratio measure can tell us *quantitatively* about *the ravens*.

In terms of the likelihood-ratio measure, and drawing on our factors p, q, and r, a reworking of Vranas's (2004) result leads to the following:

Theorem 3. *If the degree to which a non-black non-raven incrementally confirms 'All ravens are black', as measured by the likelihood-ratio, is in the interval* $1 < \mathrm{P}[\sim Ba \cdot \sim Ra | H \cdot K]/\mathrm{P}[\sim Ba \cdot \sim Ra | \sim H \cdot K] \leqslant 1 + \varepsilon$, *for very small* $\varepsilon > 0$, *then* $([\mathrm{q} - (1 - \mathrm{p})]/\mathrm{q}) < \mathrm{P}[\sim Ba | H \cdot K]/\mathrm{P}[\sim Ba | \sim H \cdot K] \leqslant ([\mathrm{q} - (1 - \mathrm{p})]/\mathrm{q}) \cdot (1 + \varepsilon)$.

If instead $(1 - \varepsilon) < \mathrm{P}[\sim Ba \cdot \sim Ra | H \cdot K]/\mathrm{P}[\sim Ba \cdot \sim Ra | \sim H \cdot K] \leqslant 1$, then $([\mathrm{q}-(1-\mathrm{p})]/\mathrm{q}) \cdot (1-\varepsilon) < (\mathrm{P}[\sim Ba | H \cdot K]/\mathrm{P}[\sim Ba | \sim H \cdot K]) \leqslant ([\mathrm{q}-(1-\mathrm{p})]/\mathrm{q})$. In both cases, for large q, $([\mathrm{q} - (1 - \mathrm{p})]/\mathrm{q}) \approx 1$, so $\mathrm{P}[\sim Ba | H \cdot K]/\mathrm{P}[\sim Ba | \sim H \cdot K] \approx 1$.[23] (Recall that $\mathrm{q} = \mathrm{P}[\sim Ba | \sim H \cdot K]/\mathrm{P}[Ra | \sim H \cdot K]$, which is plausibly quite large.)

So, the approximate independence of *Ba* from the truth or falsehood of H, given K, is a necessary condition for a non-black non-raven to provide only a very small amount of positive (or negative) support for 'All ravens are black'. Vranas's point is that traditional Bayesian treatments of the ravens paradox almost always employ the "small positive confirmation from non-black non-ravens" idea, and they inevitably draw *directly* on some such independence assumption to achieve it. But, Vranas argues, no plausible justification for *assuming* this (near) independence has yet been given by those who employ it.

Our approach sidesteps this issue completely. None of our results have relied on assuming approximate independence; indeed, our results haven't even supposed that non-black non-ravens should yield positive confirmation for H, either small or large. We've only given sufficient (and necessary) conditions for a black raven to *confirm H more* than would a non-black non-raven.

In order to address the ravens in quantitative terms, let's consider the sizes of $\mathrm{r} = \mathrm{P}[\sim Ba | H \cdot K]/\mathrm{P}[Ra | H \cdot K]$ and of $\mathrm{q} = \mathrm{P}[\sim Ba | \sim H \cdot K]/\mathrm{P}[Ra | \sim H \cdot K]$. Given background K that reflects how all of us generally believe our world to be, both r and q should presumably be quite large, and should be very nearly the same size. However, notice that such suppositions about r and q, even if we take q to precisely equal r, don't imply the approximately independence of either *Ra* or of *Ba* from H or from $\sim H$ (given K).[24]

Under such circumstances, let's consider *how much more* a black raven confirms 'All ravens are black' than does a non-black non-raven.

Theorem 4. *Given Non-triviality, suppose* $\mathrm{P}[\sim Ba | H \cdot K]/\mathrm{P}[Ra | H \cdot K] \geq \mathrm{L} > 1$ *(i.e.* $\mathrm{r} \geq \mathrm{L} > 1$*); and suppose* $\mathrm{P}[\sim Ba | \sim H \cdot K]/\mathrm{P}[Ra | \sim H \cdot K]$ *is very nearly the same* *size as r – i.e., for some* $\delta > 0$ *but near 0,* $0 < 1-\delta \leqslant (\mathrm{P}[\sim Ba|\sim H\cdot K]/\mathrm{P}[Ra|\sim H\cdot K])/(\mathrm{P}[\sim Ba|H\cdot K]/\mathrm{P}[Ra|H\cdot K]) \leqslant 1+\delta$ *(that is,* $1-\delta \leqslant \mathrm{q/r} \leqslant 1+\delta$*). Then the "ratio of likelihood ratios" is bounded as follows:*

(4.1) $$[(1-\delta)-(1-\mathrm{p})/\mathrm{L}]\cdot(1/\mathrm{p}) < \frac{\mathrm{P}[Ba\cdot Ra|H\cdot K]/\mathrm{P}[Ba\cdot Ra|\sim H\cdot K]}{\mathrm{P}[\sim Ba\cdot\sim Ra|H\cdot K]/\mathrm{P}[\sim Ba\cdot\sim Ra|\sim H\cdot K]} < (1+\delta)\cdot(1/\mathrm{p})$$

If in addition $\mathrm{P}[Ba|Ra\cdot\sim H\cdot K] > 1/2$*, then we get an improved lower bound:* $\mathrm{P}[Ba\cdot Ra|H\cdot K]/\mathrm{P}[Ba\cdot Ra|\sim H\cdot K]$

(4.2) $$(1-\delta)(1/\mathrm{p}) - 1/\mathrm{L} < \frac{\mathrm{P}[Ba\cdot Ra|H\cdot K]/\mathrm{P}[Ba\cdot Ra|\sim H\cdot K]}{\mathrm{P}[\sim Ba\cdot\sim Ra|H\cdot K]/\mathrm{P}[\sim Ba\cdot\sim Ra|\sim H\cdot K]} < (1+\delta)\cdot(1/\mathrm{p})$$

In either case, for large very $\mathrm{L} > 1$ and positive δ near 0 the "ratio of likelihood ratios" is almost exactly equal to (1/p).[25]

The larger r is, and the closer the size of q is to the size of r (i.e. the smaller δ is), the closer will be the "ratio of likelihood ratios" to 1/p. And, if instead of being nearly the same size as r, q is significantly larger than r, then q/r is significantly larger than 1 and (according to Theorem 1) the "ratio of likelihood ratios" must nearly be $(\mathrm{q/r})\cdot(1/\mathrm{p})$ (precisely $([\mathrm{q}-(1-\mathrm{p})]/\mathrm{r})\cdot(1/\mathrm{p})$), which must be significantly larger than 1/p.

Let's illustrate this theorem by plug in some fairly realistic numbers. Suppose, as seems plausible, that r is at least as large as $\mathrm{L} = 10^9$. (L should really be much larger than this since, given $H\cdot K$, it seems highly probable that there will be trillions of times more non-black things than ravens, not just billions of times more). And suppose that q is very nearly the same size as r – say, within a million of r, $\mathrm{q} = \mathrm{r} \pm 10^6$, so that $\mathrm{q/r} = 1 \pm 10^{-3}$. Then Theorem 4 tells us that for $\mathrm{P}[Ba|Ra\cdot\sim H\cdot K] = \mathrm{p} > 1/2$, the "ratio of likelihood-ratios" is bounded below by $(1-10^{-3})(1/\mathrm{p}) - 1/10^9 = (.999)(1/\mathrm{p}) - 10^{-9}$; and the upper bound is $(1+10^{-3})\cdot(1/\mathrm{p}) = (1.001)(1/\mathrm{p})$. Thus, to three significant figures the "ratio of likelihood ratios is $(1/\mathrm{p}) \pm (.001)/\mathrm{p}$.

Suppose $\mathrm{P}[Ba|Ra\cdot\sim H\cdot K] = \mathrm{p}$ is somewhere around.9 or.95; so (1/p) is somewhere around $1/.9 \approx 1.11$ or $1/.95 \approx 1.05$. Then a single instance of a black raven may not seem to yield a whole lot more support for H than a single instance of a non-black non-raven. However, under plausible conditions a sequence of n instances (i.e. of n black ravens, as compared to n non-black non-ravens) will yield a "ratio of likelihood-ratios" on the order of $(1/\mathrm{p})^{\mathrm{n}}$, which blows up significantly for large n. For example, for $\mathrm{n} = 100$ instances, $(1/.95)^{100} \approx 169$, and $(1/.9)^{100} \approx 37{,}649$ – that is, for $\mathrm{p} = .95$, 100 black raven instances would yield a likelihood-ratio 169 times higher than would 100 instances of non-black non-ravens.

Nothing in the previous paragraphs draws on the assumption that a non-black non-raven yields (at most) a tiny amount of support for H – i.e. that $\mathrm{P}[\sim Ba\cdot\sim Ra|H\cdot K]/\mathrm{P}[\sim Ba\cdot\sim Ra|\sim H\cdot K] = (1\pm\varepsilon)$. But this may be a plausible enough *additional supposition*. When it holds we have the following result.

Theorem 5. *Suppose Non-triviality, and suppose that r is large and q is very nearly the same size as r in the sense that* $(1-\delta) \leqslant \mathrm{q/r} \leqslant (1+\delta)$*, for very small* δ

(i.e., suppose the conditions for Theorem 4 hold). And suppose, in addition, that the support for H *by a non-black non-raven is very small – i.e.* $1-\varepsilon \leqslant \mathrm{P}[{\sim}Ba \cdot {\sim}Ra|H \cdot K]/\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K] \leqslant 1+\varepsilon$ *for very small* ε*. Then the support for* H *by a black raven must be*

$$\mathrm{P}[Ba \cdot Ra|H \cdot K]/\mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K] = (1 \pm \delta) \cdot (1 \pm \varepsilon) \cdot (\mathrm{P}[Ba|Ra \cdot H \cdot K]/\mathrm{P}[Ba|Ra \cdot {\sim}H \cdot K]) \approx 1/\mathrm{p}, \text{ where, of course, } \mathrm{P}[Ba|Ra \cdot H \cdot K]/\mathrm{P}[Ba|Ra \cdot {\sim}H \cdot K] = 1/\mathrm{p}.$$ [26]

Notice that the suppositions of this theorem permit a non-black non-raven to provide absolutely no support for $H (\varepsilon = 0)$, or a tiny bit of positive support ($\varepsilon > 0$), or to even provide a tiny bit of evidence against ($\varepsilon < 0$). Here, rather than *assuming* the near probabilistic independence of Ra and Ba from H and ${\sim}H$ (given K), we've effectively gotten it for free (via Theorem 3), as a consequence of the more plausible direct supposition that non-black non-ravens don't confirm much, if at all. This shows how the effect of *near independence* is accommodated by our analysis, if it happens to be implied by some *additional* plausible supposition – e.g. the assessment that no more than a minute amount of confirmation could come from an observation of a single non-black non-raven instance.

Thus, under quite weak, but highly plausible suppositions, a black raven *favors* 'All ravens are black' *more* than would a non-black non-raven by about (1/p) – i.e., by about the amount that a black object supports 'All ravens are black', given that it is a raven, since

$$\mathrm{P}[Ba|Ra \cdot H \cdot K]/\mathrm{P}[Ba|Ra \cdot {\sim}H \cdot K] = 1/\mathrm{p}.$$ [27]

This quantitative result, together with the qualitative results of the previous section, shows that a careful Bayesian analysis puts the *paradox* of the ravens to rest.

Appendix: Proofs of Various Results

Claim 1. *Given the Non-trivality Assumptions,*

$$\frac{\mathrm{P}\left[Ba \cdot Ra|H \cdot K\right]/\mathrm{P}\left[Ba \cdot Ra|{\sim}H \cdot K\right]}{\mathrm{P}\left[{\sim}Ba \cdot {\sim}Ra|H \cdot K\right]/\mathrm{P}\left[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K\right]} > 1$$

just in case $\mathrm{P}[H|Ba \cdot Ra \cdot K] > \mathrm{P}[H|{\sim}Ba \cdot {\sim}Ra \cdot K]$.

Proof. Assuming *Non-triviality* we have: $\mathrm{P}[H|Ba \cdot Ra \cdot K] > \mathrm{P}[H|{\sim}Ba \cdot {\sim}Ra \cdot K]$ *iff* both

$\mathrm{P}[H|Ba \cdot Ra \cdot K] > \mathrm{P}[H|{\sim}Ba \cdot {\sim}Ra \cdot K]$ and $\mathrm{P}[{\sim}H|Ba \cdot Ra \cdot K]$
$< \mathrm{P}[{\sim}H|{\sim}Ba \cdot {\sim}Ra \cdot K]$ *iff* $\mathrm{P}[H|Ba \cdot Ra \cdot K]/\mathrm{P}[{\sim}H|Ba \cdot Ra \cdot K]$
$> \mathrm{P}[H|{\sim}Ba \cdot {\sim}Ra \cdot K]/\mathrm{P}[{\sim}H|{\sim}Ba \cdot {\sim}Ra \cdot K]$
iff $(\mathrm{P}[Ba \cdot Ra|H \cdot K]/\mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K]) \cdot (\mathrm{P}[H|K]/\mathrm{P}[{\sim}H|K])$

$> (\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|H \cdot K]/\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K]) \cdot (\mathrm{P}[H|K]/P[{\sim}H|K])$
iff $(\mathrm{P}[Ba \cdot Ra|H \cdot K]/\mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K])/$
$(\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|H \cdot K]/\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K]) > 1.$

The following lemma establishes that all of the terms used to define p, q, and r are non-zero.

Lemma 1. *Given Non-triviality, it follows that* $\mathrm{P}[Ra|H \cdot K] > 0, \mathrm{P}[{\sim}Ba|H \cdot K]>0$, $1 > \mathrm{P}[Ra|{\sim}H \cdot K] > 0, 1 > \mathrm{P}[{\sim}Ba|{\sim}H \cdot K] > 0$, *and* $1 > \mathrm{P}[Ba|Ra \cdot {\sim}H \cdot K] > 0$.

Proof. From *Non-triviality* we have:

(i) $0 < \mathrm{P}[H|Ba \cdot Ra \cdot K] = \mathrm{P}[Ba \cdot Ra|H \cdot K] \cdot \mathrm{P}[H|K]/\mathrm{P}[Ba \cdot Ra|K]$, so $\mathrm{P}[Ra|H \cdot K] = \mathrm{P}[Ba \cdot Ra|H \cdot K] > 0$;
(ii) $0 < \mathrm{P}[H|{\sim}Ba \cdot {\sim}Ra \cdot K] = \mathrm{P}[{\sim}Ba \cdot {\sim}Ra|H \cdot K] \cdot \mathrm{P}[H|K]/\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|K]$, so $\mathrm{P}[{\sim}Ba|H \cdot K] = \mathrm{P}[{\sim}Ba \cdot {\sim}Ra|H \cdot K] > 0$;
(iii) $0 < \mathrm{P}[{\sim}H|Ba \cdot Ra \cdot K] = \mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K] \cdot \mathrm{P}[{\sim}H|K]/\mathrm{P}[Ba \cdot Ra|K]$, so $\mathrm{P}[Ba|{\sim}H \cdot K] \geq \mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K] > 0$ and $\mathrm{P}[Ra|{\sim}H \cdot K] \geq \mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K] > 0$ and $\mathrm{P}[Ba|Ra \cdot {\sim}H \cdot K] > 0$;
(iv) $0 < \mathrm{P}[{\sim}H|{\sim}Ba \cdot {\sim}Ra \cdot K] = \mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K] \cdot \mathrm{P}[{\sim}H|K]/\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|K]$, so $\mathrm{P}[{\sim}Ba|{\sim}H \cdot K] \geq \mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K] > 0$ and $\mathrm{P}[{\sim}Ra|{\sim}H \cdot K] \geq \mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K] > 0$.
(v) $0 < \mathrm{P}[{\sim}Ba \cdot Ra|K] = \mathrm{P}[{\sim}Ba \cdot Ra|H \cdot K] \cdot \mathrm{P}[H|K] + \mathrm{P}[{\sim}Ba \cdot Ra|{\sim}H \cdot K] \cdot \mathrm{P}[{\sim}H|K] = \mathrm{P}[{\sim}Ba \cdot Ra|{\sim}H \cdot K] \cdot \mathrm{P}[{\sim}H|K] < \mathrm{P}[{\sim}Ba \cdot Ra|{\sim}H \cdot K] \leqslant \mathrm{P}[Ra|{\sim}H \cdot K]$, so $0 < \mathrm{P}[{\sim}Ba|Ra \cdot {\sim}H \cdot K]$, so $\mathrm{P}[Ba|Ra \cdot {\sim}H \cdot K] < 1$.

The next claim shows how *positive support* for H depends on p and q. Our solution of *the ravens* will not depend on H receiving *positive support* (as can be seen by comparing this claim to the main Theorem, which will come next). But it's useful and interesting to see what *positive support* requires.

Claim 2. $\mathrm{P}[Ba \cdot Ra|H \cdot K]/\mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K] > 1$ *(i.e. H is positively supported by $(Ba \cdot Ra)$) if and only if* $\mathrm{P}[Ra|H \cdot K]/\mathrm{P}[Ra|{\sim}H \cdot K] > \mathrm{p}$ *(where* $\mathrm{p} = \mathrm{P}[Ba|Ra \cdot {\sim}H \cdot K]$*); and*

$\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|H \cdot K]/\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K] > 1$
(i.e. H is positively supported by$({\sim}Ba \cdot {\sim}Ra)$)
if and only if $\mathrm{P}[{\sim}Ba|H \cdot K]/\mathrm{P}[{\sim}Ba|{\sim}H \cdot K] > [\mathrm{q} - (1 - \mathrm{p})]/\mathrm{q}$.

Proof. $\mathrm{P}[Ba \cdot Ra|H \cdot K]/\mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K] = \mathrm{P}[Ra|H \cdot K]/(\mathrm{P}[Ba|Ra \cdot {\sim}H \cdot K] \cdot \mathrm{P}[Ra|{\sim}H \cdot K])$

$= (1/\mathrm{p}) \cdot \mathrm{P}[Ra|H \cdot K]/P[Ra|{\sim}H \cdot K]$
> 1 iff $\mathrm{P}[Ra|H \cdot K]/P[Ra|{\sim}H \cdot K] > \mathrm{p}$.
$\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|H \cdot K]/\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K]$

$$= \mathrm{P}[{\sim}Ba|H \cdot K]/(\mathrm{P}[{\sim}Ba|{\sim}H \cdot K] - \mathrm{P}[{\sim}Ba \cdot Ra|{\sim}H \cdot K])$$
$$= \mathrm{P}[{\sim}Ba|H \cdot K]/(\mathrm{P}[{\sim}Ba|{\sim}H \cdot K] - (1-\mathrm{p}) \cdot \mathrm{P}[Ra|{\sim}H \cdot K])$$
$$= (\mathrm{P}[{\sim}Ba|H \cdot K]/\mathrm{P}[Ra|{\sim}H \cdot K])/(\mathrm{q} - (1-\mathrm{p}))$$
$$= [\mathrm{q}/(\mathrm{q} - (1-\mathrm{p}))] \cdot \mathrm{P}[{\sim}Ba|H \cdot K]/\mathrm{P}[{\sim}Ba|{\sim}H \cdot K]$$
$$> 1 \text{ iff } \mathrm{P}[{\sim}Ba|H \cdot K]/\mathrm{P}[{\sim}Ba|{\sim}H \cdot K] > [\mathrm{q} - (1-\mathrm{p})]/\mathrm{q}.$$

Now we prove Theorem 1. We'll prove it in terms of two distinct names, 'a' and 'b', where 'a' is taken to be an instance of a black raven and 'b' is taken to be an instance of a non-black non-raven. We do it this way to assure the careful reader that no funny-business is going on when, in the main text, we treat only a single instance 'a' to see how its turning out to be a black raven compares with its turning out to be a non-black non-raven. To proceed with the treatment in terms of two possibly distinct instances we'll just need to suppose the following:

$\mathrm{P}[Bb|H \cdot K] = \mathrm{P}[Ba|H \cdot K]$, $\mathrm{P}[Rb|H \cdot K] = \mathrm{P}[Ra|H \cdot K]$, $\mathrm{P}[Bb|{\sim}H \cdot K] = \mathrm{P}[Ba|{\sim}H \cdot K]$, $\mathrm{P}[Rb|{\sim}H \cdot K] = \mathrm{P}[Ra|{\sim}H \cdot K]$, and $\mathrm{P}[Bb|Rb \cdot {\sim}H \cdot K] = \mathrm{P}[Ba|Ra \cdot {\sim}H \cdot K]$. The idea is that we have no special knowledge about b that permits us to treat it probabilistically any differently than a (prior to actually observing it). When a and b are the same instance, as in the main text, these equalities are tautological.

Theorem 1. *Given Non-triviality,* $\mathrm{q} > (1-\mathrm{p}) > 0$ *and*

$$\frac{\mathrm{P}\left[Ba \cdot Ra|H \cdot K\right]/\mathrm{P}\left[Ba \cdot Ra|{\sim}H \cdot K\right]}{\mathrm{P}\left[{\sim}Bb \cdot {\sim}Rb|H \cdot K\right]/\mathrm{P}\left[{\sim}Bb \cdot {\sim}Rb|{\sim}H \cdot K\right]} = [\mathrm{q} - (1-\mathrm{p})]/(\mathrm{p} \cdot \mathrm{r}) > 0$$

Proof. To see that $\mathrm{q} > (1-\mathrm{p})$, just observe that $\mathrm{q} = \mathrm{P}[{\sim}Ba|{\sim}H \cdot K]/\mathrm{P}[Ra|{\sim}H \cdot K] = (\mathrm{P}[{\sim}Ba \cdot Ra|{\sim}H \cdot K] + \mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K])/\mathrm{P}[Ra|{\sim}H \cdot K] = \mathrm{P}[{\sim}Ba|Ra \cdot {\sim}H \cdot K] + (\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K]/\mathrm{P}[Ra|{\sim}H \cdot K]) > (1-\mathrm{p})$, since *Non-triviality* implies $\mathrm{P}[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K] > 0$,

Non-triviality also implies (via Lemma 1) $\mathrm{p} = \mathrm{P}[Ba|Ra \cdot {\sim}H \cdot K] < 1$; so $0 < 1-\mathrm{p}$.

To get the main formula, observe that

$$(\mathrm{P}[Ba \cdot Ra|H \cdot K]/\mathrm{P}[Ba \cdot Ra|{\sim}H \cdot K])/(\mathrm{P}[{\sim}Bb \cdot {\sim}Rb|H \cdot K]/\mathrm{P}[{\sim}Bb \cdot {\sim}Rb|{\sim}H \cdot K]) = (\mathrm{P}[Ra|H \cdot K]/\{\mathrm{P}[Ra|{\sim}H \cdot K] \cdot \mathrm{p}\})/(\mathrm{P}[{\sim}Bb|H \cdot K]/\{\mathrm{P}[{\sim}Bb|{\sim}H \cdot K] - \mathrm{P}[{\sim}Bb \cdot Rb|{\sim}H \cdot K]\})$$
$$= (1/\mathrm{p})(\mathrm{P}[Ra|H \cdot K]/\mathrm{P}[Ra|{\sim}H \cdot K]) \cdot (\mathrm{P}[{\sim}Bb|{\sim}H \cdot K] - \mathrm{P}[Rb|{\sim}H \cdot K] \cdot (1-\mathrm{p}))/\mathrm{P}[{\sim}Bb|H \cdot K] = (1/\mathrm{p}) \cdot (\mathrm{P}[Ra|H \cdot K]/\mathrm{P}[Ra|{\sim}H \cdot K]) \cdot (q - (1-\mathrm{p})) \cdot (\mathrm{P}[Rb|{\sim}H \cdot K]/\mathrm{P}[{\sim}Bb|H \cdot K]) = (1/\mathrm{p}) \cdot (\mathrm{q} - (1-\mathrm{p})) \cdot (\mathrm{P}[Ra|H \cdot K]/\mathrm{P}[{\sim}Bb|H \cdot K]) \cdot (\mathrm{P}[Rb|{\sim}H \cdot K]/\mathrm{P}[Ra|{\sim}H \cdot K]) = (1/\mathrm{p}) \cdot (\mathrm{q} - (1-\mathrm{p}))/\mathrm{r}.$$

Corollary 1. *Given Non-triviality,*

$$\frac{\mathrm{P}\,[Ba \cdot Ra|H \cdot K]\,/\mathrm{P}\,[Ba \cdot Ra|{\sim}H \cdot K]}{\mathrm{P}\,[{\sim}Bb \cdot {\sim}Rb|H \cdot K]\,/\mathrm{P}\,[{\sim}Bb \cdot {\sim}Rb|{\sim}H \cdot K]} > 1 \text{ if and only if } \mathrm{q}-(1-\mathrm{p}) > \mathrm{p}\cdot\mathrm{r}.$$

And, more generally, for any real number s,

$$\frac{\mathrm{P}\,[Ba \cdot Ra|H \cdot K]\,/\mathrm{P}\,[Ba \cdot Ra|{\sim}H \cdot K]}{\mathrm{P}\,[{\sim}Ba \cdot {\sim}Ra|H \cdot K]\,/\mathrm{P}\,[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K]} = \mathrm{s}$$
$$= [\mathrm{q}-(1-\mathrm{p})]\,/\,(\mathrm{p}\cdot\mathrm{r}) > 1 \text{ if and only if } [\mathrm{q}-(1-\mathrm{p})] = \mathrm{s}\cdot\mathrm{p}\cdot\mathrm{r} > \mathrm{p}\cdot\mathrm{r}.$$

Proof. The first biconditional follows from Theorem 1 together with the obvious point that

$$[\mathrm{q}-(1-\mathrm{p})]/(\mathrm{p}\cdot\mathrm{r}) > 1 \text{ iff } \mathrm{q}-(1-\mathrm{p}) > \mathrm{p}\cdot\mathrm{r}.$$

To get the second biconditional just observe that (for any real number s),

$$\mathrm{s} = [\mathrm{q}-(1-\mathrm{p})]/(\mathrm{p}\cdot\mathrm{r}) > 1 \text{ iff } \mathrm{s}\cdot\mathrm{p}\cdot\mathrm{r} = [\mathrm{q}-(1-\mathrm{p})] > \mathrm{p}\cdot\mathrm{r}.$$

Corollary 2. *Given Non-triviality, for real number s such that*

$$\frac{\mathrm{P}\,[Ba \cdot Ra|H \cdot K]\,/\mathrm{P}\,[Ba \cdot Ra|{\sim}H \cdot K]}{[{\sim}Bb \cdot {\sim}Rb|H \cdot K]\,/\mathrm{P}\,[{\sim}Bb \cdot {\sim}Rb|{\sim}H \cdot K]} = \mathrm{s}$$
$$= [\mathrm{q}-(1-\mathrm{p})]\,/\,(\mathrm{p}\cdot\mathrm{r}),$$

we have the following*:*

(1) $\mathrm{s} > (1/\mathrm{p}) > 1$ *iff* $\mathrm{q}-(1-\mathrm{p}) > \mathrm{r}$
(2) $\mathrm{s} = (1/\mathrm{p}) > 1$ *iff* $\mathrm{q}-(1-\mathrm{p}) = \mathrm{r}$
(3) $(1/\mathrm{p}) > \mathrm{s} > 1$ *iff* $\mathrm{r} > \mathrm{q}-(1-\mathrm{p}) > \mathrm{p}\cdot\mathrm{r}$.

Proof. Follows easily from Theorem 1.

Theorem 2. *Given Non-triviality, both of the following clauses hold:*

(2.1) *If* $\mathrm{P}[{\sim}Ba|H \cdot K] > \mathrm{P}[Ra|H \cdot K]$ *(i.e. if* $\mathrm{r} > 1$*) and* $\mathrm{O}[H|Ra \cdot K]/\mathrm{O}[H|{\sim}Ba \cdot K] > (\mathrm{p} + (1-\mathrm{p}/\mathrm{r})$*, then*

$$\frac{\mathrm{P}\,[Ba \cdot Ra|H \cdot K]\,/\mathrm{P}\,[Ba \cdot Ra|{\sim}H \cdot K]}{\mathrm{P}\,[{\sim}Ba \cdot {\sim}Ra|H \cdot K]\,/\mathrm{P}\,[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K]} > 1.$$

(2.2) *If* $\mathrm{P}[{\sim}Ba|H \cdot K] \leqslant \mathrm{P}[Ra|H \cdot K]$ *(i.e.* $\mathrm{r} \leqslant 1$*), but either* $\mathrm{P}[{\sim}Ba|K] > \mathrm{P}[Ra|K]$ *or* $\mathrm{P}[{\sim}Ba\,|{\sim}H \cdot K] > \mathrm{P}[Ra\,|{\sim}H \cdot K]$ *(i.e. q > 1), then*

$$\frac{\mathrm{P}\,[Ba \cdot Ra|H \cdot K]\,/\mathrm{P}\,[Ba \cdot Ra|{\sim}H \cdot K]}{\mathrm{P}\,[{\sim}Ba \cdot {\sim}Ra|H \cdot K]\,/\mathrm{P}\,[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K]} > 1.$$

Proof. Assume Non-triviality.

Both parts of the theorem draw on the following observation:

Theorem 1 tells us that q > (1 − p) > 0 and

$$\frac{\mathrm{P}\,[Ba \cdot Ra|H \cdot K]\,/\mathrm{P}\,[Ba \cdot Ra|{\sim}H \cdot K]}{\mathrm{P}\,[{\sim}Bb \cdot {\sim}Rb|H \cdot K]\,/\mathrm{P}\,[{\sim}Bb \cdot {\sim}Rb|{\sim}H \cdot K]} = [\mathrm{q} - (1-\mathrm{p})]\,/\,(\mathrm{p} \cdot \mathrm{r})\,.$$

$$\text{So}\,\frac{\mathrm{P}\,[Ba \cdot Ra|H \cdot K]\,/\mathrm{P}\,[Ba \cdot Ra|{\sim}H \cdot K]}{\mathrm{P}\,[{\sim}Bb \cdot {\sim}Rb|H \cdot K]\,/\mathrm{P}\,[{\sim}Bb \cdot {\sim}Rb|{\sim}H \cdot K]} > 1 \text{ iff } [\mathrm{q} - (1-\mathrm{p})]\,/\,(\mathrm{p} \cdot \mathrm{r}) > 1 \text{ iff}$$

q > p · r + (1 − p) iff q/r > (p + (1 − p)/r). We will established each of the two parts of Theorem 2 by showing that their antecedents imply q/r > (p + (1 − p)/r).

(2.1) Suppose that r > 1 and O[H|$Ra \cdot K$]/O[H|$\sim Ba \cdot K$] > (p + (1 − p)/r). Then

$$\begin{aligned}\mathrm{q/r} &= (\mathrm{P}[{\sim}Ba|{\sim}H \cdot K]/\mathrm{P}[Ra|{\sim}H \cdot K])/(\mathrm{P}[{\sim}Ba|H \cdot K]/\mathrm{P}[Ra|H \cdot K])\\ &= (\mathrm{P}[Ra|H \cdot K]/\mathrm{P}[Ra|{\sim}H \cdot K])/(\mathrm{P}[{\sim}Ba|H \cdot K]/\mathrm{P}[{\sim}Ba|{\sim}H \cdot K])\\ &= \frac{(\mathrm{P}\,[Ra|H \cdot K]\,/\mathrm{P}\,[Ra|{\sim}H \cdot K]) \cdot (\mathrm{P}\,[H|K]\,/\mathrm{P}\,[{\sim}H|K])}{(\mathrm{P}\,[{\sim}Ba|H \cdot K]\,/\mathrm{P}\,[{\sim}Ba|{\sim}H \cdot K]) \cdot (\mathrm{P}\,[H|K]\,/\mathrm{P}\,[{\sim}H|K])}\\ &= \frac{\mathrm{O}\,[H|Ra \cdot K]}{\mathrm{O}\,[H|{\sim}Ba \cdot K]} > (\mathrm{p} + (1-\mathrm{p})/\mathrm{r}).\end{aligned}$$

(2.2) Suppose P[$\sim Ba$|$H \cdot K$] ⩽ P[Ra|$H \cdot K$] (i.e. r ⩽ 1), but either P[$\sim Ba$|K] > P[Ra|K] or P[$\sim Ba$|$\sim H \cdot K$] > P[Ra|$\sim H \cdot K$] (i.e. q > 1).

First we show that we must have q > 1 in any case. This is shown by reductio, as follows:

Suppose q ⩽ 1. Then P[$\sim Ba$|K] > P[Ra|K].
So we have P[$\sim Ba$|$\sim H \cdot K$] ⩽ P[Ra|$\sim H \cdot K$] (i.e. q ⩽ 1) and
P[$\sim Ba$|$H \cdot K$] ⩽ P[Ra|$H \cdot K$] (i.e. r ⩽ 1). Then

$$\begin{aligned}&\mathrm{P}[{\sim}Ba|K] = \mathrm{P}[{\sim}Ba|H \cdot K]\mathrm{P}[H|K] + \mathrm{P}[{\sim}Ba|{\sim}H \cdot K]\mathrm{P}[{\sim}H|K]\\ &\leqslant \mathrm{P}[Ra|H \cdot K]\mathrm{P}[H|K] + \mathrm{P}[Ra|{\sim}H \cdot K]\mathrm{P}[{\sim}H|K] = \mathrm{P}[Ra|K] < \mathrm{P}[{\sim}Ba|K]\end{aligned}$$

Contradiction!!!

Thus we have q > 1 and r ⩽ 1; so1/r ≥ 1. Then (p + (1 − p)/r) ⩽ p/r + (1 − p)/r = 1/r < q/r.

Theorem 3. *If the degree to which a non-black non-raven incrementally confirms 'All ravens are black', as measured by the likelihood-ratio, is in the interval* 1 < P[$\sim Ba \cdot \sim Ra$|$H \cdot K$]/P[$\sim Ba \cdot \sim Ra$|$\sim H \cdot K$] ⩽ 1 + ε, *for very small* ε > 0, *then* ([q − (1 − p)]/q) < P[$\sim Ba$|$H \cdot K$]/P[$\sim Ba$|$\sim H \cdot K$] ⩽ ([q − (1 − p)]/q)(1 + ε).

If instead (1 − ε) < P[$\sim Ba \cdot \sim Ra$|$H \cdot K$]/P[$\sim Ba \cdot \sim Ra$|$\sim H \cdot K$] ⩽ 1, *then*

$$([\mathrm{q} - (1-\mathrm{p})]/\mathrm{q})(1-\varepsilon) < (\mathrm{P}[{\sim}Ba|H \cdot K]/\mathrm{P}[{\sim}Ba|{\sim}H \cdot K]) \leqslant ([\mathrm{q} - (1-\mathrm{p})]/\mathrm{q}).$$

In both cases, for large q, $([q-(1-p)]/q) \approx 1$, *so* $P[{\sim}Ba|H \cdot K]/P[{\sim}Ba|{\sim}H \cdot K] \approx 1$.

Proof. $P[{\sim}Ba \cdot {\sim}Ra|H \cdot K]/P[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K] = P[{\sim}Ba|H \cdot K]/(P[{\sim}Ba|{\sim}H \cdot K] - P[{\sim}Ba \cdot Ra|{\sim}H \cdot K]) = P[{\sim}Ba|H \cdot K]/(P[{\sim}Ba|{\sim}H \cdot K] - (1-p) \cdot P[Ra|{\sim}H \cdot K]) = (P[{\sim}Ba|H \cdot K]/P[Ra|{\sim}H \cdot K])/(q-(1-p)) = (P[{\sim}Ba|H \cdot K]/P[{\sim}Ba|{\sim}H \cdot K]) \cdot q/(q-(1-p))$. So, for $1 < P[{\sim}Ba \cdot {\sim}Ra|H \cdot K]/P[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K] \leqslant (1+\varepsilon)$, $([q-(1-p)]/q) < (P[{\sim}Ba|H \cdot K]/P[{\sim}Ba|{\sim}H \cdot K]) \leqslant ([q-(1-p)]/q)(1+\varepsilon)$. Also, if $(1-\varepsilon) < P[{\sim}Ba \cdot {\sim}Ra|H \cdot K]/P[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K] \leqslant 1$, then $([q-(1-p)]/q) \cdot (1-\varepsilon) < (P[{\sim}Ba|H \cdot K]/P[{\sim}Ba|{\sim}H \cdot K]) \leqslant [q-(1-p)]/q$.

Acknowledgements We would like to thank the following people for useful conversations about the paradox of confirmation: Luc Bovens, Fabrizio Cariani, Kenny Easwaran, Ted Hailperin, Alan Hájek, Stephan Hartmann, Chris Hitchcock, Colin Howson, Franz Huber, Jim Joyce, Patrick Maher, Chad Mohler, Brad Monton, Mike Titelbaum, and Brian Weatherson. Special thanks to Jan Sprenger, whose critique of the implications of an earlier version of the results in Sections "A New Bayesian Approach to the Paradox" and "Quantitative Results" spurred us to make significant improvements.

Notes

[1] For a nice taste of this voluminous literature, see the bibliography in Vranas (2004).

[2] Almost all early commentators on the paradox have viewed (EC) and premise (2) as beyond reproach. But not all contemporary commentators are so sanguine about (EC) and (2). See Sylvan and Nola (1991) for detailed discussion of non-classical logics and the paradoxes of confirmation. See Gemes (1999) for a probabilistic approach that also denies premise (2). We will not discuss such approaches here. We restrict our focus to accounts couched in terms of classical logic.

[3] Interestingly, while Hempel and Goodman are completely unsympathetic to Quine's strategy *here*, they are much more sympathetic to such maneuvers in the context of the *Grue* Paradox. In this sense, Quine's approach to the paradoxes is more unified and systematic than Hempel's or Goodman's, since they give "special treatment" to Grue-predicates, while Quine views the problem – in *both* paradoxes of confirmation – to be rooted in the "non-naturalness" of the referents of the predicates involved. For what it's worth, we think a unified and systematic approach to the paradoxes is to be preferred. But, we think a unified *Bayesian* approach is preferable to Quine's instantial approach. However, our preferred Bayesian treatment of Grue will have to wait for another paper.

[4] Perhaps Hempel had something like the following in mind. Notice that $(\forall x)(Rx \supset Bx)$ entails *Ba* given *Ra*; so, given *Ra*, ~*Ba* falsifies $(\forall x)(Rx \supset Bx)$ and, on Hempel's account, *Ba* confirms it. Likewise, $(\forall x)(Rx \supset Bx)$ entails ~*Ra* given ~*Ba*; so, given ~*Ba*, *Ra* falsifies $(\forall x)(Rx \supset Bx)$ and, on Hempel's account, ~*Ra* confirms it. However, $(\forall x)(Rx \supset Bx)$ entails neither *Ba* nor ~*Ba* given ~*Ra*; so, arguably, one might hold that $(\forall x)(Rx \supset Bx)$ cannot be confirmed by either *Ba* or by ~*Ba* given ~*Ra* (though, as already affirmed, it is confirmed by ~*Ra* given ~*Ba*). Similarly, $(\forall x)(Rx \supset Bx)$ entails neither *Ra* nor ~*Ra* given *Ba*; so, arguably, one might hold that $(\forall x)(Rx \supset Bx)$ cannot be confirmed by either *Ra* or by ~*Ra* given *Ba* (though, of course, it is confirmed by *Ba* given *Ra*). Even if a Hempelian story along these lines can be told, it won't save Hempel's analysis from problem #2, below.

[5] Hypothetico-deductive approaches to confirmation also imply (M), since they explicate "E confirms H relative to K" as "H & K entails E." So, H-D confirmation cannot avail itself of a Hempel-style resolution of the paradox either.

[6] Maher notes that Hempel *never proves* that (PC) is true while (PC*) is false. This is an understatement! He *cannot* prove this claim, *on pain of contradiction with his official theory of confirmation*. We think the reason Hempel (and others) missed this inconsistency is that it is easy to conflate *objectual* and *propositional* senses of "confirmation". If you think of the *objects* doing the confirming, then one can see (PC) as true and (PC*) as false (even from a deductivist point of view). But, if you think of the *propositions* as doing the confirming, then this is impossible from a deductivist point of view (i.e., from the point of view of any theory which entails (M)). The salient passages from Hempel suggest that he slides back and forth between objectual and propositional senses of confirmation. And, we suspect that this is what led him into the present inconsistency.

[7] We won't bother to discuss the competing axiomatizations and interpretations of probability. These details won't matter for our discussion. For simplicity we will just assume that P is some rational credence function, and that it satisfies an appropriate version of the standard (Kolmogorov 1956) axioms. But these assumptions could be altered in various ways without affecting the main points we will make below.

[8] Metaphysically, there may be a problem with "non-natural kinds" (in Quine's sense - e.g., disjunctive and negative properties) participating in certain kinds of causal or other law-like relations. This sort of problem has been suggested in the contemporary literature by Armstrong (1978), Shoemaker (1980), and others. But, we think this metaphysical fact (if it is a fact) has few (if any) *confirmational* consequences. Confirmation is a logical or epistemic relation, which may or may not align neatly with metaphysical relations like causation or law-likeness.

[9] As Chihara (1981) points out, "there is no such thing as *the* Bayesian solution. There are many different 'solutions' that Bayesians have put forward using Bayesian techniques". That said, we present here what we take to be the most standard assumptions Bayesians tend to make in their handling of the paradox - assumptions that are *sufficient* for the desired comparative and quantitative confirmation-theoretic claims. On this score, we follow Vranas (2004). However, not all Bayesians make precisely these assumptions. To get a sense of the variety of Bayesian approaches, see, *e.g.*: (Alexander 1958); (Chihara 1981); (Earman 1992); (Eells 1982); (Gaifman 1979); (Gibson 1969); (Good 1960, 1961); (Hesse 1974); (Hooker & Stove 1968); (Horwich 1982), (Hosiasson-Lindenbaum 1940); (Howson & Urbach 1993); (Jardine 1965); (Mackie 1963); (Nerlich 1964); (Suppes 1966); (Swinburne 1971, 1973); (Wilson 1964); (Woodward 1985); (Hint ikka 1969); (Humburg 1986); (Maher 1999, 2004); (Vranas 2004).

[10] We take *logarithms* of the ratio measures just to ensure that they are positive in cases of confirmation, negative in cases of disconfirmation, and zero in cases of neutrality of irrelevance. This is a useful convention for present purposes, but since *log*s don't alter the ordinal structure of the measures, it is a *mere* convention.

[11] This has led some former defenders of s to abandon it as a measure of incremental confirmation. See Joyce (2004, fn. 11). See, also, Eells and Fitelson (2000, 2002) and Fitelson (2001) for further peculiarities of the measure s.

[12] Often, Bayesians use a *two-stage sampling* model in which *two* objects a and b are sampled at random from the universe, and where K_α entails $(Ra \cdot {\sim}Bb)$ (e.g., Earman 1992). On that model we still have (2), but (3) is replaced with $P[{\sim}Bb|H \cdot K_\alpha] = P[{\sim}Bb|K_\alpha]$, and (COMP$_P$) is replaced by (COMP$'_P$)$P[H|Ra \cdot Ba \cdot K_\alpha] > P[H|{\sim}Bb \cdot {\sim}Rb \cdot K_\alpha]$. However, no real loss of generality comes from restricting our treatment to "one-stage sampling" – i.e., to the selection of a single object a, which K_α doesn't specify to be either an R or a ${\sim}B$ (Vranas 2004, fns. 10 and 18). We prefer a *one*-stage sampling approach because it simplifies the analysis somewhat, and because we think it is closer in spirit to what Hempel and Goodman took the original paradox to be about – where K_α is assumed *not* to have any implications about the color or species of the objects sampled, and where a single object is observed "simultaneously" for its color and species.

[13] However, Vranas does not argue that (3′) is false or implausible – only that no good argument for its plausibility has been given. So, it is consistent with his result that one might be able to find some plausible condition X that, together with (1′), implies (QUANT$_c$). Vranas' result would then show that condition X (together with (1′)) also implies (3′) – and so in effect would

provide a plausibility argument for (3′). Some of the results we prove in the next two sections will provide such conditions, X.

[14] $P[H|Ba \cdot Ra \cdot K] > P[H|{\sim}Ba \cdot {\sim}Ra \cdot K]$ iff $P[{\sim}H|Ba \cdot Ra \cdot K] < P[{\sim}H|{\sim}Ba \cdot {\sim}Ra \cdot K]$. So, $P[H|Ba \cdot Ra \cdot K] > P[H|{\sim}Ba \cdot {\sim}Ra \cdot K]$ iff $P[H|Ba \cdot Ra \cdot K]/P[{\sim}H|Ba \cdot Ra \cdot K] > P[H|{\sim}Ba \cdot {\sim}Ra \cdot K]/P[{\sim}H|{\sim}Ba \cdot {\sim}Ra \cdot K]$ iff $P[Ba \cdot Ra|H \cdot K]/P[Ba \cdot Ra|{\sim}H \cdot K] > P[{\sim}Ba \cdot {\sim}Ra|H \cdot K]/P[{\sim}Ba \cdot {\sim}Ra|{\sim}H \cdot K]$.

[15] Throughout the remainder of the paper our treatment will focus on the relationship between these likelihood-ratios. However, for $0 < P[H|K] < 1$, we have $P[H|Ba{\cdot}Ra{\cdot}K] > P[H|{\sim}Ba{\cdot}{\sim}Ra{\cdot}K]$ if and only if $c[H, Ba \cdot Ra|K] > c[H, {\sim}Ba \cdot {\sim}Ra|K]$, where c is any of the three measures of incremental confirmation d, r, and l. This is the result (†) discussed in the previous section, together with its (easy to established) converse. So, a specific qualitative relationship (>, or =, or <) holds between these likelihood-ratios just in case it holds between $P[H|Ba \cdot Ra \cdot K]$ and $P[H|{\sim}Ba \cdot {\sim}Ra \cdot K]$, just in case it holds between $c[H, Ba \cdot Ra|K]$ and $c[H, {\sim}Ba \cdot {\sim}Ra|K]$, where c is any of the measures d, r, and l.

[16] That is, the conditions we will establish do not imply that likelihood-ratio $P[Ba{\cdot}Ra|H \cdot K]/P[Ba{\cdot}Ra|{\sim}H \cdot K]$ is itself greater than 1. And, since this likelihood-ratio will be greater than 1 *just when* H receives positive support from $(Ba \cdot Ra)$ (i.e. *just when* $P[H|Ba \cdot Ra \cdot K] > P[H|K]$), it follows that we will not be requiring that H receive positive support from $(Ba \cdot Ra)$. (See Claim 2 in the Appendix for more about this.)

[17] $P[Ba \cdot Ra|K] > 0$ and $P[{\sim}Ba \cdot {\sim}Ra|K] > 0$ are required for $P[H|Ba \cdot Ra \cdot K]$ and $P[H|{\sim}Ba \cdot {\sim}Ra \cdot K]$ to be well-defined; $0 < P[H|Ba \cdot Ra \cdot K] < 1$ implies $0 < P[H|K] < 1$. Other implication of *Non-trivality* are in Appendix Lemma 1.

[18] Since for $q = r$, $s > 1$ iff $q - (1 - p) > p \cdot q$ iff $q(1 - p) = q - pq > (1 - p)$ iff $q > 1$.

[19] One clause of this result draws on the notion of *odds*, O. By definition, $O[X|Y] = P[X|Y]/P[{\sim}X|Y]$.

[20] An equivalent (and perhaps more illuminating) alternative to the second condition for Clause (2.1) is this: the ratio $P[Ra|H \cdot K]/P[Ra|{\sim}H \cdot K]$ is no less than the ratio $P[{\sim}Ba|H \cdot K]/P[{\sim}Ba|{\sim}H \cdot K]$, or perhaps only a bit less – i.e. $(P[Ra|H \cdot K]/P[Ra|{\sim}H \cdot K])/(P[{\sim}Ba|H \cdot K]/P[{\sim}Ba|{\sim}H \cdot K]) \geq (p + (1 - p)/r)$. Here $p + (1 - p)/r < 1$ because the first condition of Clause 2.1 requires $r > 1$. This condition (and the equivalent odds condition) is *strictly weaker* than the usual independence assumptions. For, if independence assumption (2) holds, then the $P[Ra|H \cdot K]/P[Ra|{\sim}H \cdot K] = 1$, and if independence assumption (3) holds, then the $P[{\sim}Ba|H \cdot K]/P[{\sim}Ba|{\sim}H \cdot K] = 1$. Thus, the two usual conditions entail the much more restrictive $P[Ra|H \cdot K]/P[Ra|{\sim}H \cdot K] = P[{\sim}Ba|H \cdot K]/P[{\sim}Ba|{\sim}H \cdot K]$ – i.e. $O[H|Ra \cdot K] = O[H|{\sim}Ba \cdot K]$.

[21] Vranas (2004) provides a detailed exposition.

[22] We'll suppress the "log", since nothing we'll say depends on the re-scaling of likelihood-ratios by taking the log.

[23] This approximate independence condition implies approximate independence condition (1′), since $P[{\sim}Ba|K] = P[{\sim}Ba|H \cdot K] \cdot P[H|K] + P[{\sim}Ba|{\sim}H \cdot K] \cdot (1 - P[H|K]) \approx P[{\sim}Ba|H \cdot K] \cdot P[H|K] + P[{\sim}Ba|H \cdot K] \cdot (1 - P[H|K]) = P[{\sim}Ba|H \cdot K]$. The two versions of approximate independence are equivalent if $P[H|K]$ isn't extremely close to 1.

[24] To see this clearly, supposes that $P[Ra|H \cdot K]$ is larger than $P[Ra|{\sim}H \cdot K]$ by a very large factor $f > 1$ – i.e. $P[Ra|H \cdot K] = f \cdot P[Ra|{\sim}H \cdot K]$ – and suppose that $P[{\sim}Ba|H \cdot K]$ is larger than $P[{\sim}Ba|{\sim}H \cdot K]$ by the same factor – i.e. $P[{\sim}Ba|H \cdot K] = f \cdot P[{\sim}Ba|{\sim}H \cdot K]$. Then we'd have $r = P[{\sim}Ba|H \cdot K]/P[Ra|H \cdot K] = P[{\sim}Ba|{\sim}H \cdot K]/P[Ra|{\sim}H \cdot K] = q$ even though neither Ra nor Ba would be anywhere close to independence of H or ${\sim}H$. The same goes for $P[Ra|{\sim}H \cdot K]$ larger than $P[Ra|H \cdot K]$ and $P[{\sim}Ba|{\sim}H \cdot K]$ larger than $P[{\sim}Ba|H \cdot K]$, both by very large factor $f > 1$.

[25] Proof: $(P[Ba{\cdot}Ra|H \cdot K]/P[Ba{\cdot}Ra|{\sim}H \cdot K])/(P[{\sim}Ba{\cdot}{\sim}Ra|H \cdot K]/P[{\sim}Ba{\cdot}{\sim}Ra|{\sim}H \cdot K]) = (q/r)(1/p) - [(1 - p)/p]/r$, by Theorem 1. We get the upper bounds as follow: $(q/r)(1/p) - [(1 - p)/p]/r < (q/r)(1/p) \leqslant (1 + \delta) \cdot (1/p)$. To get the lower bound in (4.1): $(q/r) \cdot (1/p) - [(1 - p)/p]/r > (1 - \delta)/p - 1/pr \geq [(1 - \delta) - 1/L] \cdot (1/p)$. To get the lower bound in (4.2), first

notice that for $p > 1/2$, $[(1-p)/p] < 1$, so $(q/r)\cdot(1/p)-[(1-p)/p]/r \geq (1-\delta)\cdot(1/p)-1/r > (1-\delta)\cdot(1/p)-1/L$.

[26] Proof: From Theorem 3 we already have that $P[\sim Ba\cdot\sim Ra|H\cdot K]/P[\sim Ba\cdot\sim Ra|\sim H\cdot K] = (1\pm\varepsilon)$ implies $P[\sim Ba|H\cdot K]/P[\sim Ba|\sim H\cdot K] = (1\pm\varepsilon)$. Then $P[Ra|H\cdot K]/P[Ra|\sim H\cdot K] = [(P[Ra|H\cdot K]/P[Ra|\sim H\cdot K])/(P[\sim Ba|H\cdot K]/P[\sim Ba|\sim H\cdot K])]\cdot(P[\sim Ba|H\cdot K]/P[\sim Ba|\sim H\cdot K]) = (q/r)\cdot(1\pm\varepsilon) = (1\pm\delta)(1\pm\varepsilon)$. And $P[Ba\cdot Ra|H\cdot K]/P[Ba\cdot Ra|\sim H\cdot K] = (P[Ba|Ra\cdot H\cdot K]/P[Ba|Ra\cdot\sim H\cdot K])(P[Ra|H\cdot K]/P[Ra|\sim H\cdot K])$.

[27] The factor $p = P[Ba|Ra\cdot\sim H\cdot K]$ is a reflection of both likelihoods and prior probabilities for the whole range of alternative hypotheses H_f, where each says that the frequency of black things among ravens, F[Bx, Rx] = f, is a specific fraction f. When p is pretty close to 1, the only alternative hypotheses H_f that can have non-miniscule prior probabilities are those for which f is pretty close to 1 as well. So a single black raven doesn't provide very much confirmation for H (i.e., only about 1/p, which isn't much), because it takes a lot of instances to distinguish between H and the alternatives that have f near 1. To see this formally, consider: for each $k \geq 1$ such that $1/(1-p) > k$, $p = P[Ba|Ra\cdot\sim H\cdot K] = \Sigma_{1>f\geq 0}P[Ba|Ra\cdot H_f\cdot K]P[H_f|Ra\cdot\sim H\cdot K] = \Sigma_{1>f\geq 0}f\cdot P[H_f|Ra\cdot\sim H\cdot K] < \Sigma_{1>f\geq 1-k(1-p)}P[H_f|Ra\cdot\sim H\cdot K] + [1-k\cdot(1-p)]\Sigma_{1-k(1-p)>f\geq 0}P[H_f|Ra\cdot\sim H\cdot K] = 1-k\cdot(1-p)\cdot\Sigma_{1-k(1-p)>f\geq 0}P[H_f|Ra\cdot\sim H\cdot K]$, so $\Sigma_{1-k\cdot(1-p)>f\geq 0}P[H_f|Ra\cdot\sim H\cdot K] < 1/k$; thus $\Sigma_{1>f\geq 1-k(1-p)}P[H_f|Ra\cdot\sim H\cdot K] > (k-1)/k$. For example, for $p = .98$ and $k = 5$ we have that $\Sigma_{1\geq f>.90}P[H_f|Ra\cdot\sim H\cdot K] > .80$. Indeed, in this case one only gets $\Sigma_{1\geq f>.90}P[H_f|Ra\cdot K]$ to be as small as .80 by making $P[H_{.9}|Ra\cdot\sim H\cdot K] = .20$ and $P[H_{.999}|Ra\cdot\sim H\cdot K] = .80$, and $P[H_f|Ra\cdot K] = 0$ for all other values of f. If non-zero priors are more evenly distributed throughout the interval between.9 and 1, then $\Sigma_{1\geq f>.90}P[H_f|Ra\cdot\sim H\cdot K]$ has to be quite a bit larger than .80 (in order to permit $p = .98$).

References

Alexander HG (1958) The paradoxes of confirmation. Br J Philos Sci 9:227–233

Armstrong D (1978) A theory of universals. Cambridge University Press, Cambridge

Carnap R (1950) Logical foundations of probability, 2nd edn. Chicago University Press, Chicago, 1962

Carnap R (1952) The continuum of inductive methods. University of Chicago Press, Chicago, IL

Carnap R (1971) A basic system of inductive logic, part I. In: Carnap R, Jeffrey RC (eds) Studies in inductive logic and probability, vol 1. University of California Press, Berkeley, CA

Carnap R (1980) A basic system of inductive logic, Part II. In Jeffrey RC (ed) Studies in Inductive Logic and Probability, vol 2. University of California Press, Berkeley, CA

Chihara C (1981) Quine and the confirmational paradoxes. In French PA, Uehling Jr. TE, Wettstein HK (eds) Midwest studies in philosophy, vol. 6. The foundations of analytic philosophy. University of Minnesota Press, Minneapolis, pp 425–452

Earman J (1992) Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory. MIT Press, Cambridge, MA

Eells E (1982) Rational decision and causality. Cambridge University Press, New York

Eells E, Fitelson B (2000) Measuring confirmation and evidence. J Philos XCVII:663–672

Eells E, Fitelson B (2002) Symmetries and asymmetries in evidential support. Philos Stud 107:129–142

Fitelson B (1999) The plurality of bayesian measures of confirmation and the problem of measure sensitivity. Philos Sci 66:S362–S378

Fitelson B (2001) Studies in Bayesian confirmation theory. Ph.D. dissertation, University of Wisconsin

Fitelson B (2004) Inductive logic forthcoming. In: Pfeife J, Sarkar S (eds) The Philosophy of science: an encyclopedia. Routledge, London

Gaifman H (1979) Subjective probability, natural predicates and Hempel's Ravens. Erkenntnis 14:105–147

Gemes K (1999) Carnap-confirmation, content-cutting, and real confirmation. unpublished essay, Yale

Gibson L (1969) On 'Ravens and Relevance' and a likelihood solution of the paradox of confirmation. Br J Philos Sci 20:75–80

Good IJ (1960) The paradox of confirmation. Br J Philos Sci 11:145–149

Good IJ (1961) The paradox of confirmation (II). Br J Philos Sci 12:63–64

Good IJ (1967) The white shoe is a red herring. Br J Philos Sci 17:322

Good IJ (1968) The white shoe *qua* herring is pink. Br J Philos Sci 19:156–157

Goodman N (1954) Fact, Fiction, and Forecast. Athlone Press, London

Hempel CG (1945) Studies in the logic of confirmation I & II. Mind 54

Hempel CG (1967) The white shoe: no red herring. Br J Philos Sci 18:239–240

Hesse M (1974) The Structure of Scientific Inference. MacMillan, London

Hintikka J (1969) Inductive independence and the paradoxes of confirmation. In: Rescher N (ed) Essays in honor of Carl G. Hempel: a tribute on the occasion of his sixty-fifth birthday. Reidel, Dordrecht, pp 24–46

Hooker CA, Stove D (1968) Relevance and the ravens. Br J Philos Sci 18:305–315

Horwich P (1982) Probability and evidence. Cambridge University Press, New York

Hosiasson-Lindenbaum J (1940) On confirmation. J Symb Logic 5:133–148

Howson C, Urbach P (1993) Scientific reasoning: The Bayesian approach, 2nd edn. Open Court, Chicago, IL

Humburg J (1986) The solution of Hempel's raven paradox in Rudolf Carnap's system of inductive logic. Erkenntnis 24:57–72

Jardine R (1965) The resolution of the confirmation paradox. Aus J Philos 43:359–368

Joyce J (Winter 2003 Edition) Bayes's theorem. In: Zalta EN (ed) The stanford encyclopedia of philosophy. http://plato.stanford.edu/archives/win2003/entries/bayes-theorem/

Kolmogorov A (1956) Foundations of probability, 2nd English edn. AMS Chelsea Publishing, Providence, RI

Mackie JL (1963) The paradox of confirmation. Br J Philos Sci 13:265–277

Maher P (1999) Inductive logic and the ravens paradox. Philos Sci 66:50–70

Maher P (2004) Probability captures the logic of scientific confirmation. In: Hitchcock C (ed) Contemporary debates in the philosophy of science. Blackwell, Oxford, pp 69–93

Nerlich G (1964) Mr. Wilson on the paradox of confirmation. Aus J Philos 42:401–405

Quine WVO (1969) Natural kinds. In: Quine WVO (ed) Ontological relativity and other essays. Columbia University Press, New York, pp 114–138

Shoemaker S (1980) Causality and properties. Reprinted in: Identity, cause and mind. Cambridge University Press, Cambridge

Suppes P (1966) A Bayesian approach to the paradoxes of confirmation. In: Hintikka J, Suppes P (eds) Aspects of inductive logic. North-Holland, Amsterdam, pp 198–207

Swinburne R (1971) The paradoxes of confirmation – a survey. Am Philos Quart 8:318–330

Sylvan R, Nola R (1991) Confirmation without paradoxes. In: Schurz G, Dorn GJW (eds) Advances in scientific philosophy: essays in honour of Paul Weingartner. Rodopi, Amsterdam, pp 5–44

Vranas P (2004) Hempel's Raven paradox: a lacuna in the standard Bayesian solution. Br J Philos Sci 55:545–560

Wilson PR (1964) A new approach to the confirmation paradox. Aus J Philos 42:393–401

Woodward J (1985) Critical review: Horwich on the ravens, projectability and induction. Philos Stud 47:409–428

Learning to Network

Brian Skyrms and Robin Pemantle

Introduction

In species capable of learning, including our own, individuals can modify their behavior by some adaptive process. Important classes of behavior – mating, predation, coalitions, trade, signaling, and division of labor – involve interactions between individuals. The agents involved learn two things: with *whom to interact* and *how to act*. That is to say that adaptive dynamics operates both on structure and strategy.

In an interaction, individuals actualize some behavior, the behavior of the individuals jointly determines the outcome of the interaction, and the consequences for the individuals motivate learning. At this high level of abstraction, we can model interactions as games. The relevant behaviors of individuals are called strategies of the game, and the strategies of the players jointly determine their payoffs. Payoffs drive the learning dynamics (Skyrms and Pemantle 2000).

If we fix the interaction structure in this abstract scheme, we get models of the evolution of strategies in games played on a fixed structure. An interaction structure need not be deterministic. In general, it can be thought of as a specification of the probabilities of interaction with other individuals. By far the most frequently studied interaction structure is one in which the group of individuals is large and individuals interact at random. That is to say that each individual has equal probability of interacting with every other individual in the population. Among a list of virtues of this model, mathematical tractability must come near the top. At another end of the spectrum we have models where individuals interact with their neighbors on a torus, or a circle, or (less frequently) some other graphical structure (Ellison 1993; Nowak and May 1992; Hegselmann 1996; Alexander 2000). Except in the simplest cases, these models sacrifice mathematical tractability to gain realism, and

B. Skyrms (✉)
Logic and Philosophy of Science, University of California, Irvine, CA 92697
and
Department of Philosophy, Stanford University, Stanford, CA 94305
e-mail: bskyrms@uci.edu

R. Pemantle
Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104
e-mail: pemantle@math.upenn.edu

E. Eells and J.H. Fetzer (eds.), *The Place of Probability in Science*, Boston Studies in the Philosophy of Science 284, DOI 10.1007/978-90-481-3615-5_12,

computer simulations have played an important role in their investigation. These two extreme models, however, can have quite different implications for the evolution of behavior. In large, random encounter settings cooperators are quickly eliminated in interactions with a Prisoner's Dilemma structure. Comparable local interaction models allow cooperators to persist in the population.

If we fix the strategies of individuals and let the interaction structure evolve we get a model of interaction network formation. Evolution of structure is less well-studied than evolution of strategies, and is the main focus of this paper. Most current research on theory of network formation takes the point of view that networks are modeled as graphs or directed graphs, and network dynamics consists of making and breaking of links (Jackson and Watts 2002; Bala and Goyal 2000). In taking an interaction structure to be a specification of probabilities of interaction rather than a graphical structure, we take a more general view than most of the literature (but see Kirman 1997 for a point of view close to that taken here). It is possible that learning dynamics may drive these probabilities to zero or one and that a deterministic graphical interaction structure may crystallize out, but this will be treated as a special case. We believe that this probabilistic approach can give a more faithful account of both human and non-human interactions. It also makes available a set of mathematical tools that do not fit the coarser picture of making or breaking deterministic links in a graphical structure.

Ultimate interest resides in the general case where structure and strategy coevolve. These may be modified by the same or different kinds of learning. They may proceed at the same rate or different rates. The case where structure dynamics is slow and strategy dynamics is fast may approximate more familiar models where strategies evolve on a fixed interaction structure. The opposite case may be close to that of individuals with fixed strategies (or phenotypes) learning to network. In between, there is a very rich territory waiting to be explored. We will close this paper with a discussion of the co-evolution of structure and strategy in a game which one of us has argued is the best simple prototype of the problem of instituting a social contract (Skyrms 2004). Whether coevolution of structure and strategy supports or reverses the conventional wisdom about equilibrium selection in this game, depends on the nature and relative rates of the two learning processes.

Learning

Learning can be divided into two broad categories: (1) belief learning in which the organism forms beliefs (or internal representations of the world) and uses these to make decisions, and (2) reinforcement learning, where the organism increases the probability of acts that have been rewarded and decreases the probability of those that have not been rewarded. Ultimately the distinction may not be so clear cut, but it is useful for a categorization of learning theories. In the simplest belief learning model, Cournot dynamics, an individual assumes that others will do what they did last time and performs the act that has the highest payoff on that assumption. More sophisticated individuals might form their beliefs more carefully, by applying inductive reasoning to some or all of the available evidence. Less confident individuals might hedge their bet on Cournot dynamics with some probabilistic version of the rule. Strategically minded individuals might predict the effect of their current choice on future choices of the other agents involved, and factor this into their decision. Humans, having a very large brain, can do all of these things but often they do not bother (Suppes and Atkinson 1960; Roth and Erev 1995; Erev and Roth 1998; Busemeyer and Stout 2002; Yechiam and Busemeyer 2005).

Reinforcement learning does not require a lot of effort, or a large brain, or any brain at all. In this paper we will concentrate on reinforcement learning, although we will also touch on other forms of learning. Specifically, we apply a mathematical model in which the probability of an act is proportional to the accumulated rewards from performing that act Herrnstein 1970; Roth and Erev 1995). Following Luce (1959), the learning model can be decomposed into two parts: (i) a *reinforcement dynamics*, in which weights or propensities for acts evolve, and (ii) a *response rule*, which translates these weights into probabilities of acts. If we let weights evolve by adding the payoff gotten to the weight of the act chosen, and let our probabilities be equal to the normalized weights (Luce's linear response rule), we get the basic Herrnstein–Roth–Erev dynamics.

There are alternative models of reinforcement learning that could be investigated in this setting. In a path-breaking study, Suppes and Atkinson (1960) applied stimulus sampling dynamics to learning in two-person games. Borgers and Sarin (1997) have investigated dynamics of Bush and Mosteller (1955) in a game-theoretic setting. Instead of the Luce's linear response rule of normalizing the weights, some models use a logistic response rule. Bonacich and Liggett (2004) apply Bush–Mosteller learning in a setting closely resembling our own. They get limiting results that are closely connected to this of the discounted model of Friends II in Skyrms and Pemantle (2000). Liggett and Rolles (2004) generalize the results of Bonacich and Liggett to an infinite space of agents. We, however, will concentrate attention on the basic Herrnstein–Roth–Erev dynamics and on a "slight" variation on it.

Erev and Roth (1997) suggest modifying the basic model by discounting the past to take account of "forgetting". At each time period, accumulated weights are multiplied by some positive discount factor less than one, while new reinforcements are added at full strength. Discounting is a robust phenomenon in experimental studies of reinforcement learning, but there seems to be a great deal of individual variability with reported discount factors ranging from .5 to .99 (Erev and Roth 1997; Busemeyer and Stout 2002; and Busemeyer 2004; Goeree and Holt forthcoming). Discounting changes the limiting properties of the learning process radically. We will see that within the reported range of individual variability, small variations in the discount rate can lead to large differences in predicted observable outcomes in interactive learning situations.

Two-Person Games with Basic Reinforcement Learning

We begin by investigating basic (undiscounted) reinforcement learning in simple two-person interactions. The following model was introduced in Skyrms and Pemantle (2000). Each day each individual in a small group wakes up and decides to visit someone. She decides by chance, with the chance of visiting anyone else in the group being given by normalized weights for that individual. (We can imagine the process starting with some initial weights; they can all be set to one to start the process with random encounters.) The person selected always accepts, and there is always time enough in the day for all selected interactions to take place. In a group of ten, if Jane decides to visit someone and the other nine all happen to decide to visit Jane, she has a total of ten interactions in that day. Each interaction produces a payoff. At the end of the day, each individual updates her weights for every other individual by adding the payoffs gotten that day from interactions with that individual. (Obvious variations on the basic model suggest themselves, but we confine ourselves here to just this model applied to different kinds of interactions.) Initially, we investigate baseline cases where individuals have only the choice of with whom to interact, and interactions always produce payoffs in the same way. Then we build on the results for these cases to analyze interactions in the stag hunt game, in which different agents can have different acts and the combination of acts determines the payoffs.

Consider two games of "Making Friends." In Friends I the visitor is always treated well, and gains a payoff of 1, while the host goes to some trouble but also enjoys the encounter, for a net payoff of zero. In Friends II the visitor and host are both equally reinforced, with a payoff of 1 going to each. We start each learning process with each individual having initial weights of one for each other individual, so that our group begins by interacting at random. It is easy to run computer simulations of the Friends I and Friends II processes, and it is a striking feature of such simulations that in both cases non-random interaction structure rapidly emerges. Furthermore, rerunning the processes from the same starting point seems to generate different structure each time. In this setting, we should expect the emergence of structure without an organizer, or even an explanation in terms of payoff differences. The state of uniform random encounters with which we started the system does not persist, and so must count as a very artificial state. Its use as the fixed interaction structure in many game theoretic models is therefore extremely suspect.

We can understand the behavior of the Friends I process if we notice that each individual's learning process is equivalent to a Polya urn. We can think of him as having an urn with balls of different colors, one color for each other individual. Initially there is one ball of each color. A ball is chosen (and returned), the designated individual is visited. Because visitors are always reinforced, another ball of the same color is added to the urn. Because only visitors are reinforced, balls are not added to the urn in any other way. (Philosophers of science will be familiar with the Polya urn because of its equivalence with Bayes–Laplace inductive inference.) The Polya urn converges to a limit with probability one, but it is a random limit with uniform distribution over possible final probabilities. Anything can happen, and nothing is favored! In Friends I the random limit is uniform for each player, and makes the players independent (Skyrms and Pemantle 2000, Theorem 1). All interaction structures are possible in the limit, and the probability that the group converges to random encounters is zero.

In Friends II, both visitor and host are reinforced and so the urns interact. If someone visits you, you are reinforced to visit him – or to put it graphically, someone can walk up to your door and put a ball of his color in your urn. This complicates the analysis. Nevertheless, the final picture is quite similar. The limiting probabilities must be *symmetric*, that is to say X visits Y with the same probability that Y visits X, but subject to this constraint and its consequences anything can happen (Skyrms and Pemantle 2000; Theorem 2).

So far, the theory has explained the surprising results of the simulations, but a rather special case of Friends II provides a cautionary contrast. Suppose that there are only three individuals. (What we are about to describe is much less likely to happen if the number of individuals is a little larger.) Then the only way we can have symmetric visiting probabilities is if each individual visits the other two each with probability one-half. Then the previous theorem implies that in this case the process must converge to these probabilities. In simulations this sometimes happens rapidly. However, there are other trials in which the system appears to be converging to a state in which individual A visits B and C equally, but B and C always visit A and never each other. You can think of individual A as "Ms. Popular." The system was observed to stay near such a state for a long time (5,000,000 iterations of the process.)

This apparent contradiction is resolved in Pemantle and Skyrms (2004b), using the theory of stochastic approximation. For the basic Herrnstein–Roth–Erev model, there is an underlying deterministic dynamics that can be obtained from the expected increments of the stochastic process. This deterministic dynamics has four equilibria – one in which each individual visits the others with equal probability and the other three having A, B, and C respectively as "Ms. Popular." The symmetric equilibrium is strongly stable – an attractor – while the "Ms. Popular" equilibria are unstable saddle points. The system must converge to the symmetric equilibrium. It cannot converge to one of the unstable saddles, but if in the initial stages of learning it wanders near a saddle it may take a long time to escape because the vector pushing it away is very small. This is what happens in the anomalous simulations. There is a methodological moral here that we will revisit in the next section. Simulations may not be a reliable guide to limiting behavior and limiting behavior is not necessarily all that is of interest.

The Making Friends games provide building blocks for analyzing learning dynamics where the interactions are games with non-trivial strategies. Consider the two-person Stag Hunt. Individuals are either Stag Hunters or Hare Hunters. If a Stag Hunter interacts with a Hare Hunter no Stag is caught and the Stag Hunter gets zero payoff. If a Stag Hunter interacts with another Stag Hunter the Stag is likely caught and the hunters each get a payoff of one. Hare Hunting requires no cooperation, and its practitioners get a payoff of .75 in any case. The game is of special interest for social theory because cooperation is both mutually beneficial and an equilibrium, but it is risky (Skyrms 2004). In game theoretic terminology, Stag hunting is payoff dominant and Hare hunting is risk dominant. In a large population composed of half Stag Hunters and half Hare Hunters with random interactions between individuals, the Hare Hunters would get an average payoff of .75 while the Stag Hunters would only get an average payoff of .50. The conventional wisdom is that in the long run evolution will strongly favor Hare hunting, but we say that one should consider the possibility that the players *learn to network*.

We use exactly the same model as before, except that the payoffs are now determined by the individuals' types or strategies: Hunt Stag or Hunt Hare. We start with an even number of Stag Hunters and Hare Hunters. Theory predicts that, in the limit, Stag Hunters always visit Stag Hunters and Hare Hunters always visit Hare Hunters (Skyrms and Pemantle 2000; Th. 6). Simulation confirms that such a state is approached rapidly. Although on rational choice grounds Hare Hunters "should not care" whom they visit, they cease to be reinforced by visits from Stag Hunters after Stag Hunters learn not to visit them. Hare Hunters continue to be visited by other Hare Hunters, so all the differential learning for Hare Hunters takes place when they are hosts rather than visitors. Once learning has sorted out Stag Hunters and Hare Hunters so that each group only interacts with its own members, each is playing Friends II with itself and previous results characterize within-group interaction structure.

Now Stag Hunters prosper. Was it implausible to think that Stag Hunters might find a way to get together? If they were sophisticated, well-informed, optimizing agents they would have gotten together right away! Our point is that it doesn't take much for Stag Hunters to get together. A little bit of reinforcement learning is enough.

Clique Formation with Discounting the Past

Adding a little discounting of the past is a natural and seemingly modest modification of the reinforcement process. However, it drastically alters the limiting behavior of learning. If the Polya urn, which we used in the analysis of Friends I, is modified by discounting the past the limiting result is that after some time (can't say when) there will be one color (can't say which) that will always be picked. Discounting the past, no matter how little the discounting, leads to deterministic outcomes. This is also true when we learn to network. Discounting the past leads to the formation of cliques, whose members never interact with members of alternative cliques. Why then, did we even bother to study learning without discounting? We will see that if discounting is small enough, learning with discounting may, for long periods of time, behave like learning without discounting.

The effects of adding discounting to the learning process are already apparent in two-person interactions (Skyrms and Pemantle 2000), but they are more interesting in multi-person interactions. Here we discuss two three-person interactions, Three's Company (a uniform reinforcement counterpart to Friends II), and a Three-Person version of the Stag Hunt. Every day, each individual picks two other individuals to visit to have a three-person interaction. The probability of picking a pair of individuals is taken to be proportional to the product of their weights. The payoff that an individual receives from a three-person interaction is added to her weights for each of the other two participants. We again start the learning process with random interaction. Everyone begins having weight one for everyone else.

In Three's Company, as in Friends II, everyone is always reinforced in every interaction. Everyone gets a payoff of one. No matter what the discount rate, the limiting result of discounted learning is clique formation. For a population size of six or more, the population will break up into cliques of size 3, 4, or 5. Each member of a given clique chooses each other member of that clique with positive limiting relative frequency. For each member of a clique, there is a finite time after which she does not choose outsiders. All such cliques – that is each partition of the population into sets of size 3, 4, and 5, has positive probability of occurring (Pemantle and Skyrms 2004b, Th. 4.1).

Simulations at a discount rate of .5 conform to theory. A population of 6 always broke into two cliques of size 3, with no interactions between cliques. As we discount less – keeping more of the past weights – we see a rapid shift in results. Multiplying past weights by .6, led to formation of two cliques in 994/1000 trials; by .7 in 13/1000; by .8 in none. (We ran the process for 1,000,000 time steps and rounded interaction probabilities to two decimal places.) Writing the discount factor by which past payoffs are multiplied as $(1 - x)$, we can say that simulation says that clique formation occurs reliably for large x, but not at all for small x with a large transition taking place between $x = .4$ and $x = .3$. The theory says that clique formation occurs for any positive x.

This apparent conflict between theory and simulation is resolved in Pemantle and Skyrms (2004a), where it is shown that time to clique formation increases exponentially in $1/x$ as the discount factor $(1 - x)$ approaches 1. The behavior of the process for observable finite sequences of iterations is highly sensitive to the discount parameter, within ranges that fall within the individual variability that has been reported in the experimental literature. When x is close to 1, discounted reinforcement learning behaves for long periods of time like undiscounted learning in which clique formation almost never occurs.

Three's Company, like Friends II, is important because it arises naturally in the analysis of less trivial interactions. Consider a Three-Player Stag Hunt (Pemantle and Skyrms 2004a). Pairs of individuals are chosen, and weights evolve, just as in Three's Company, but the payoffs depend on the types of players. If three Stag Hunters interact, they all get a payoff of 4, but a Stag Hunter who has at least one Hare Hunter in his trio gets nothing. (In a random encounter setting, Stag Hunting is here even more risky that in the two person case.) Hare hunters always get a payoff of 3.

In the limit Stag Hunters learn to always visit other Stag Hunters but, unlike some other limiting results we have discussed, this one is attained very rapidly. With 6 Stag Hunters and 6 Hare Hunters and a discount rate of .5, the probability that a stag hunter will visit a hare hunter usually drops below half a percent in 25 interactions. In 50 iterations this always happened in 1,000 trials, and this remains true for values of x between .5 and .1. For $x = .01$, 100 iterations suffices and 200 iterations are enough if $x = .001$.

Once Stag Hunters learn to visit Stag Hunters, they are essentially playing a game of Three's Company among themselves. They may be visited by Hare Hunters, but these visits produce no reinforcement for the Stag Hunters and so do not alter their weights. Stag Hunters then form cliques of size 3, 4, or 5 among themselves. This will take a long time if the past is only slightly discounted.

There is a tendency for Hare Hunters to learn to visit Hare Hunters after the Stag Hunters learn not to visit them, but because of the discounting it is possible for a Hare Hunter to be frozen in a state of visiting one or two Stag Hunters. This is a real possibility when the past is heavily discounted. At $x = .5$, at least one Hare Hunter interacted with a Stag Hunter (after 10,000 iterations) in 384 out of 1,000 trials. This dropped to 6/1,000 for $x = .2$ and to 0 for $x = .1$. Hare Hunters who are no trapped into interactions with Stag Hunters eventually end up playing Three's Company among themselves and also form cliques of size 3, 4, and 5.

Coevolution of Structure and Strategy

So far we have concentrated on the dynamics of interaction, because we believe that it has not received as much attention as it deserves. The full story involves co-evolution of both interaction structure and strategy. Depending on the application, these may involve the same or different adaptive dynamics and they may evolve at the same or different rates. We will illustrate this with two different treatments of the two-person Stag Hunt.

To the two-person Stag Hunt of the section Two-Person Games with Basic Reinforcement Learning, we add a strategy revision process based on imitation. This *reinforcement-imitation* model was discussed in Skyrms and Pemantle (2000). With some specified probability, an individual wakes up, looks around the whole group, and if some strategy is prospering more than his own, switches to it. Individual's probabilities are independent. If imitation is fast relative to structure dynamics, it operates while individuals interact more or less at random and Hare Hunters will take over more often than not. If imitation is slow, stag hunters find each other and prosper, and then imitation slowly converts Hare Hunters to Stag Hunters (who quickly learn to interact with other Stag Hunters).

Simulations show that in intermediate cases, timing can make all the difference. We start with structure weights equal to 1 and vary the relative rates of the dynamics by varying the imitation probability. With "fast" imitation ($pr = .1$) 78% of the trials ended up with everyone converted to Hare Hunting and 22% ended up with everyone converted to Stag Hunting. Slower imitation ($pr = .01$) almost reversed the numbers, with 71% of the trials ending up All Stag and 29% ending up All Hare. Fluid network structure coupled with slow strategy revision reverses the orthodox prediction that Hare Hunting (the risk dominant equilibrium) will take over.

(This conclusion remains unaffected if we add discounting to the learning dynamics for interaction structure. Discounting the past simply means that Stag

Hunters find each other more rapidly. No matter how Hare Hunters end up, Stag Hunters are more prosperous. Imitation converts Hare Hunters to Stag Hunters.).

The foregoing model illustrates the combined action of two different dynamics, reinforcement learning for interaction structure and imitation for strategy revision. *What happens if both processes are driven by reinforcement learning?* In particular, we would like to know whether the relative rates of structure and strategy dynamics still make the same difference between Stag Hunting and Hare Hunting. In this *Double Reinforcement* model, each individual has two weight vectors, one for interaction propensities and one for propensities to either Hunt Stag or Hunt Hare. Probabilities for whom to visit and what to do are both gotten by normalizing the appropriate weights. Weights are updated by adding the payoff from an interaction to both the weight for the individual involved and to the weight for the action taken. Relative rates of the two learning processes can be manipulated by changing the magnitude of the initial weights.

In the previous models we started the population off with some Stag Hunters and some Hare Hunters. That point of view is no longer correct. The only way one could be deterministically a Stag Hunter would be if he started out with zero weight for Hare Hunting, and then he could never learn to hunt Stag. We have to start out individuals with varying propensities to hunt Hare and Stag. There are various interesting choices that might be made here; we will report some simulation results for one. We start with a group of 10, with 2 confirmed Stag Hunters (weight 100 for Stag, 1 for Hare), 2 confirmed Hare Hunters (weight 100 for Hare, 1 for Stag), and 6 undecided guys (weights 1 for Stag and 1 for Hare. Initial weights for interaction structure were all equal, but their magnitude was varied from .001 to 10, in order to vary the relative rates of learning structure and strategy. The percent of 10,000 trials that ended up All Stag or All Hare (after 1,000,000 iterations) for these various settings are shown in Fig. 1. As before, fluid interaction structure and slow strategy adaptation favor Stag Hunting, while the reverse combination favors Hare Hunting.

In both reinforcement-imitation and double reinforcement models of the coevolution of structure and strategy a fluid network structure shifts the balance form the risk dominant Hare Hunting equilibrium to the cooperative Stag Hunt.

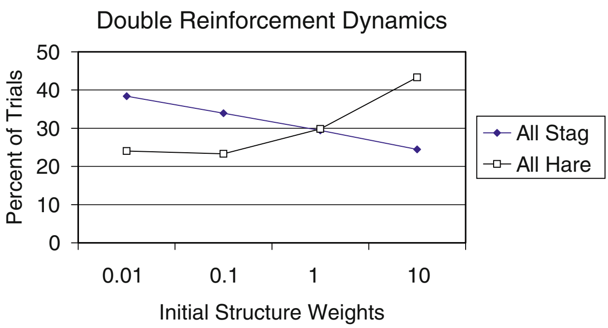


Fig. 1 Stag hunt with reinforcement dynamics for both strategy and structure

Why Dynamics?

Classical, pre-dynamic, game theory would approach the problem differently. The whole group of 10 individuals is playing a 10-person game. A move consists in choosing both a person to play with and a strategy. We can just identify the Nash equilibria of this large game. None are strict. The pure equilibria fall into two classes. One class has everyone hunting Stag and every possible interaction structure. The other has everyone hunting Hare and every possible interaction possible interaction structure. (There are also mixed equilibria with every possible interaction structure.) From this point of view, interaction structure does not seem very important. If you ignore dynamics you miss a lot.

References

Alexander JM (2000) Evolutionary explanations of distributive justice. Philos Sci 67:490–516
Bala V, Goyal S (2000) A non-cooperative model of network formation. Econometrica 68: 1181–1229
Bonacich P, Liggett T (2003) Asymptotics of a matrix-valued markov chain arising in sociology. Stochastic Processes Appl 104:155–171
Busemeyer J, Stout J (2002) A contribution of cognitive decision models to clinical assessment: decomposing performance on the bechara gambling task. Psychol Assess 14:253–262
Bush R, Mosteller F (1955) Stochastic models of learning. Wiley, New York
Ellison G (1993) Learning, local interaction, and coordination. Econometrica 61:1047–1071
Erev I, Roth A (1998) Predicting how people play games: reinforcement learning in experimental games with unique mixed strategy equilibria. Am Econ Rev 88:848–881
Goeree JK, Holt CA (2002) in Encyclopedia of Cognitive Science, Nadel L (ed) (McMillan, London) 2:1060–1069
Hegselmann R (1996) Social dilemmas in lineland and flatland. In: Liebrand W, Messick D (eds) Frontiers in social dilemmas research. Springer, Berlin, pp 337–362
Herrnstein RJ (1970) On the law of effect. J Exp Anal Behav 13:243–266
Jackson M, Watts A (2002) On the formation of interaction networks in social coordination games. Games Econ Behav 41:265–291
Kirman A (1997) The economy as an evolving network. J Evol Econ 7:339–353
Liggett TM, Rolles S (2004) An infinite stochastic model of social network formation. Stochastic Processes Appl 113:65–80
Luce RD (1959) Individual choice behavior. Wiley, New York
Nowak M, May R (1992) Evolutionary games and spatial chaos. Nature 359:826–829
Pemantle R, Skyrms B (2004a) Time to absorption in discounted reinforcement models. Stochastic Processes and Their Appl. 109:1–12
Pemantle R, Skyrms B (2004b) Network formation by reinforcement learning: the long and the medium run. Math Behav Sci 48:315–327
Pemantle R, Skyrms B (in preparation) Reinforcement Schemes May Take a Long Time to Exhibit Limiting Behavior
Roth A, Erev I (1995) Learning in extensive form games: experimental models and simple dynamic models in the intermediate term. Games Econ Behav 8:14–212
Skyrms B (2004) The stag hunt and the evolution of social structure. Cambridge, New York

Skyrms B, Pemantle R (2000) A dynamic model of social network formation. Proc Natl Acad Sci USA 97:9340–9346

Suppes P, Atkinson R (1960) Markov learning models for multiperson interactions. Stanford University Press, Palo Alto, CA

Yechiam E, Busemeyer JR (2005) Comparison of basic assumptions embedded in learning models for experienced based decision-making. Psychon Bull Rev 12:387–402

Probabilities in Decision Rules

Paul Weirich

Decision theory advances an account of rational behavior. It uses a type of probability accessible to agents to guide their decisions, usually subjective probability, that is, probability for an agent. Personalized, a probability assignment indicates an agent's degrees of belief at a time. If an agent is cognitively ideal and rational, her degrees of belief conform to the standard probability axioms and so form a type of probability.

This paper examines subjective probabilities in principles of rational decision making. The principles are important in the behavioral sciences such as economics (especially microeconomics) and psychology (especially cognitive psychology). I explain how the goal of accessibility motivates making propositions the objects of probabilities and then relativizing probabilities to ways of grasping propositions. These are ways of understanding propositions and have sentence-like structure. My points about them apply given various accounts of their nature. They may, for instance, be mental representations as in Crimmins (1992) and Fodor (1998, Chap. 1), modes of presentation as in McKinsey (1999), or epistemic intensions as in Chalmers (2002).

The proposed relativization promotes psychological realism. Personal pronouns, demonstratives, proper names, and other expressions refer directly without the intermediary of a concept. Hence two sentences may express the same proposition. An ideal agent who understands both sentences may nonetheless miss their synonymy and consequently assign two probabilities to the proposition they express. Relativization to ways of grasping propositions resolves such inconsistency. It generalizes the usual probability laws and returns them as a special case when an agent knows all relevant synonymies.

Relativization is a complication that theorists rightly are reluctant to accept. Realism in decision theory, however, compels making probabilities relative to ways of grasping propositions. An agent's probability assignment to a proposition requires her understanding the proposition. If there are two ways of understanding

P. Weirich (✉)
Department of Philosophy, University of Missouri, Columbia, MO 65211, USA
e-mail: weirichp@missouri.edu

E. Eells and J.H. Fetzer (eds.), *The Place of Probability in Science*, Boston Studies in the Philosophy of Science 284, DOI 10.1007/978-90-481-3615-5_13,

the proposition, she may assign two probabilities to the proposition. This possibility requires that a proposition's probability be relative to a way of grasping it.

Relativizing probabilities to ways of grasping propositions makes a decision's rationality relative to the way the agent frames the decision. Most framing effects arise from logical mistakes and misunderstanding propositions. However, because of multiple ways of understanding the same proposition, they may also arise from lack of empirical information. Then they affect the rationality of decisions made by cognitively ideal agents. Acknowledging this influence makes normative decision theory more realistic.

The first four sections make the case for relativization. They present standard probability and decision principles and show that relativization handles the phenomenon of direct reference. The next three sections carry out the relativization aiming for minimal revision of probability and decision principles.

Subjective Probability

How does subjective probability compare with other types of probability? Subjective probabilities, in contrast with physical probabilities, depend on evidence. The physical probability of drawing a particular ball from an urn containing several balls depends on factors such as the number of balls in the urn, whereas the subjective probability of drawing that ball depends on one's evidence concerning such factors.

Theorists debate the extent to which evidence controls rational degrees of belief and so subjective probabilities. Harsanyi (1967) and Aumann (1976) hold that rational ideal agents with the same information make the same probability assignments. They hold that evidence completely settles rational degrees of belief. Binmore (1994, pp. 61–63, 206–216) rejects that view. He holds that rationality gives agents some latitude in assigning probabilities. Morris (1995) explores the issue, which I leave open.

Physical probabilities manifest themselves in relative frequencies of results of a repeated experiment, for example, the relative frequency of Heads in repeated tosses of a coin. Taking a physical probability to be a relative frequency, the probability attaches to an event-type, with some event-type serving as a reference class. Taking the physical probability to be a propensity, it attaches to an event relative to a physical situation. For instance, the probability that a uranium atom decays in a certain time period is relative to a physical situation including time. It has one value at a time before the atom has decayed and another value at a time after the atom has decayed.

Subjective probabilities may rest on evidence concerning frequencies but are single-case probabilities. For instance, an agent's subjective probability that a particular coin toss yields Heads may rest on his evidence concerning the coin's symmetry, the tossing mechanism, and the history of results from past tosses. Still, the probability concerns the outcome of a single toss. Moreover, it is not relative to a reference class or a physical situation, although evidence concerning classes to which the toss belongs and concerning the physical situation of the toss influences the probability. Decision rules need such single-case probabilities because decisions concern action in a single-case, for example, betting on the outcome of a particular coin toss. They also need probabilities depending on evidence only, because an agent may not know pertinent physical probabilities relative to appropriate reference classes or physical situations.

To what does a subjective probability attach? In other words, what is its object? A subjective probability, being a degree of belief, has the same object as a belief, namely, a proposition. Suppose that an agent believes that Chicago is in Illinois. The object of his belief is the proposition that Chicago is in Illinois. If he holds this belief with certainty, then he assigns to that proposition a degree of belief equal to 1, the maximum degree of belief.

Propositions are intensional objects, and nuances of meaning differentiate them. The theory of rational choice benefits from making objects of probability as fine-grained as propositions are. A rational agent's certainty that Chicago is in Illinois and uncertainty that Chicago is in the land of Lincoln may motivate a difference in response to questions and wagers about Chicago's location. Hence probabilities should distinguish the attitudes. A difference in doxastic attitude that motivates a difference in behavior normally grounds a difference in objects of probability. Economic theory recognizes the value of making fine-grained the objects of attitudes that generate behavior. For instance, Debreu (1959, pp. 28–29, 32) introduces dated and located commodities in place of generic commodities to acknowledge that dates and locations influence a consumer's decisions about commodities.

Some theorists, such as Savage (1972), take subjective probabilities to attach to events, more precisely, possible events. The difference between possible events and propositions may not be large. Although an actual event taken as a concrete object differs from a true proposition taken as an abstract object, the nature of events and propositions is an open topic, and some accounts identify possible events with propositions. Nonetheless, unification of theories of belief and degrees of belief drives the prevalent account of subjective probabilities to taking propositions as their objects. Jeffrey (1983) is a standard bearer for this account.

Subjective probability is intensional in virtue of applying to intensional objects such as propositions. Some authors who discuss the intensionality of subjective probabilities are Gärdenfors (1975) and Horgan (2000). Some authors who discuss the intensionality of a decision's possible outcomes, a closely related topic, are van Fraassen (2002, pp. 103–108) and Schick (2003, Chap. 1).

Subjective probability is degree of belief in a proposition, but what is a proposition? It is a meaning of a declarative sentence and a basic bearer of a truth-value. According to some accounts, for example Stalnaker (1984), propositions are sets of possible worlds. The proposition that Chicago is in Illinois is the set of worlds in which it is true that Chicago is in Illinois. According to this view of propositions, only one necessary proposition exists. A rival view of Kaplan (1989) and others attributes to propositions structures similar to sentences' structures and distinguishes necessary propositions according to their structures. I take propositions to have structure and, moreover, to be compositional, that is, to be meanings that are a function of the meanings of their parts.

Although propositions resemble sentences, they differ from sentences because ambiguous sentences and sentences with indexicals express different propositions in different contexts. For instance, the sentence, "It is raining," is true at some times and places but false at other times and places. Also, different sentences may express the same proposition. For instance, "It is raining," and the French sentence, "Il pleut," are synonymous.

Explicating subjective probabilities in terms of degrees of belief leads to questions about the latter. The subject is complex, and I clarify just a few key points. An agent's degree of belief in a proposition may represent a range of doxastic attitudes from certainty that the proposition is false (and so not belief at all) to certainty that the proposition is true. Also, a degree of belief need not represent a quantitative mental state. It is a quantitative representation of a mental state, which may be non-quantitative, as Jeffrey (1992, p. 29) and Weirich (2004b) explain.

In some cases information is too meager to support a precise probability assignment to a proposition. Also, in some cases an agent fails to assign a proposition a probability because making the assignment is pointless. For simplicity, I explore only cases in which a rational ideal agent with a good, although incomplete, stock of information assigns a precise probability to every proposition relevant to current decision problems.

Some authors define degrees of belief operationally in terms of preferences concerning bets and support their definitions with the type of representation theorem common in measurement theory. The theorem shows that if a rational ideal agent has preferences with a rich structure, then his probability and utility assignments are inferable from his preferences. In this school, degrees of belief satisfy the probability laws by definition. Consequently, the laws are not norms for degrees of belief. Instead, they generate a norm for preferences: Have preferences over gambles that probability and utility functions may represent as agreeing with expected utilities. If an agent's preferences cannot be represented that way, then they are incoherent.

Textbooks usually present probability laws as norms for degrees of belief. For example, they say that if the probability of getting two Heads from two coin tosses is 1/4, then an agent should assign 1–1/4, or 3/4, as the probability of getting at least one Tails from those tosses. They use the probability laws to guide probability assignments. I follow the textbooks rather than operationism and take the probability laws as normative constraints on degrees of belief. Degrees of belief not satisfying the probability laws are possible, but irrational in an ideal agent. Betting behavior is evidence for, but not definitive of, degrees of belief, as Weirich (2001, Sec. 1.4) explains. A degree of belief of $x\%$ represents a propositional attitude. It is not just part of a mathematical representation of preferences among gambles.

Taking a proposition's probability as the expression of an attitude toward the proposition raises many interesting questions that operational definitions squelch. This paper treats one: Can an agent assign two probabilities to the same proposition? Taking probability doxastically, an agent needs to grasp a proposition to assign it a probability. Dispositions to bet do not suffice. Can an agent grasp a proposition two ways and assign a different probability given each way of grasping it? I address cases where this happens even though the agent is rational and ideal.

To conclude this section, let me mention and set aside two problems concerning the view that propositions are the objects of probabilities. Because the problems are resolvable, they do not refute the view.

First, taking the objects of probability as propositions generates probabilities of probabilities. These higher-order probabilities threaten inconsistency. They generate paradoxes in the style of the liar paradox, as Weirich (1983, pp. 87–89) observes. They also generate paradoxes of exhaustiveness, as Skryms (1980, pp. 109–111) notes. If every set of propositions generates a proposition, say, the proposition that all the propositions in the set are true, then there cannot be a set of all propositions. For every set's cardinality is less than the cardinality of the set of all its subsets. Standard methods of dealing with the semantic and set-theoretic paradoxes, however, handle these problems. One may introduce a hierarchy of probability assignments. Also, applications may work with a finite set of possible worlds specifying just what matters to an agent. Skryms (1984, pp. 29–36), Gaifman (1988), Sahlin (1994), and Weirich (2004b) address the issues.

Second, taking propositions as objects of probabilities invites an agent to assign probabilities to propositions expressing his possible acts. Jeffrey (1983) and Joyce (1999) acknowledge such probability assignments. On the other hand, Levi (2000) objects to them, claiming that they preclude choice. He holds that an agent choosing an option is not in a position to make predictions about his choice. Joyce (2002) responds to the objections, arguing that an agent who regards his choices as efficacious must assign probabilities to his acts.

Applications of Probability Laws

An agent has access to his own degrees of belief and so is responsible for regulating them and taking account of them. He is irrational if he lets his degrees of belief run afoul of principles governing them and makes decisions unresponsive to them. How does an agent's access to his probability assignment affect applications of standard probability and decision principles? The principles have two main interpretations. According to one interpretation, they are principles that an agent uses to guide his probability judgments and his decisions. They direct an agent's deliberations. According to another interpretation, they are standards of evaluation that an outsider may use to appraise for rationality an agent's probability judgments and decisions. A child may conform with these standards without using them to guide behavior. Under the first interpretation, an agent's access to his probability assignments clearly matters. He cannot use the principles to form probability judgments and make decisions if he lacks information their application requires. Under the second interpretation, an agent's access also matters. Suppose an agent makes probability judgments and decisions that violate the principles. An extenuating factor is his lack of access to the probabilities the violations involve. Their inaccessibility may excuse the violations. The agent may not be in a position to know about the violations and correct them. Whether a violation is inexcusable from rationality's standpoint depends in part on whether the agent commits the violation knowingly. Standards of rationality are attainable in the sense that agents can comply knowingly. Agents are thus responsible for shortcomings.

According to my view, in many although not all cases probability and decision principles furnish rational procedures. Applying the probability laws clearly requires access to the probabilities they govern. An agent uses the laws effectively only if she knows the propositions assigned as values to propositional variables and the probabilities attached to those propositions. Consider $P(s)$ for some state s. To know the probability, an agent needs knowledge of the state and the probability function, at least as applied to the state. By design, subjective probabilities are accessible. They equal an agent's own degrees of belief. Although a nonideal agent may be unaware of her degrees of belief, an ideal agent is aware of them.

Taking probability and decision principles to yield rational procedures for ideal agents clearly motivates an interpretation of probabilities that makes them accessible to those agents. It motivates taking them as degrees of belief attached to propositions. Later sections argue that it also motivates making probabilities of propositions relative to ways of grasping propositions. An agent lacks access to a probability if it attaches to a proposition independently of her grasp of the proposition. An agent's assignment of a probability to a proposition should acknowledge its dependence on her understanding of the proposition. Acknowledging that dependence creates a problem, however. An agent may grasp a proposition two ways and so assign it two probabilities. Accessibility requires a way of grasping the proposition, but ways of grasping a proposition may generate two probabilities for the same proposition. Neither the procedural nor even the evaluative interpretation of the probability and decision principles licenses addressing the problem by giving up accessibility.

The standard probability laws arise in the context of background laws concerning the existence of probabilities. For subjective probabilities, one assumes a finite set of atomic propositions to which the operations of negation, disjunction, and conjunction apply indefinitely to yield the closure of the base set under those operations. The closure constitutes an algebra of propositions, and probabilities attach to the propositions in the algebra. The standard axioms governing their probabilities are:

For all s, $P(s) \geq 0$.
For all s such that s is an a priori truth, $P(s) = 1$.
For all s and r such that s and r are inconsistent a priori, $P(s \vee r) = P(s) + P(r)$.

A theorem following from these axioms states that $P(s) = P(r)$ if s and r are equivalent a priori.

This presentation of the basic laws of probability puts aside the topic of conditional probability. That topic raises issues independent of relativization, such as the nature of conditional probabilities and the requirement to update probabilities by conditionalization. Also, it puts aside algebras formed from infinite sets of atomic propositions because they raise independent issues concerning, for instance, countable additivity.

The background laws make $P(s)$ relative to an algebra of which s is a member. Can s be atomic in two algebras and have one probability relative to one algebra and another probability relative to the other algebra? The probability axioms do not rule out this possibility, but an implicit coherence requirement does. Coherence requires that a proposition have just one probability and so the same probability relative to each algebra containing it.

Some classic examples show how various partitions of possible events, and the algebras the partitions generate, may lead to multiple probabilities for an event. The eighteenth century French probabilist d'Alembert argued that the probability of two Heads on two coin tosses is 1/3 because the possible cases are two Heads, two Tails, and a mixture of Heads and Tails. An agent under d'Alembert's spell may assign $P(HH)$ the value 1/3, dividing possible cases as he did, but also assign it the value 1/4, dividing possible cases the usual way. Consequently, the probability the agent assigns two Heads depends on how she frames the relevant possible cases.

Bertrand's paradox also draws attention to such framing effects. Drop a straw across a barrel top. What is the probability that the chord it forms has a length greater than half the barrel top's diameter? Taking lengths of the chord as equiprobable yields one answer. Taking lengths of the arc cut by the chord as equiprobable yields another answer. An agent who falls victim to such framing effects is inconsistent in the identification of equiprobable base cases. He assigns equal probabilities to the atomic propositions of rival algebras. He should not take both algebras' atomic propositions to be equiprobable. Attention to base cases, however, does not prevent inconsistent probabilities that arise from two ways of grasping the same proposition.

Necessary, logical, and a priori truths form distinct categories, which are, respectively, metaphysical, formal, and epistemic. According to the usage I follow, an a priori truth is roughly a true proposition whose truth is apparent to a fully rational ideal agent regardless of her evidence. The laws of probability attend to a priori features of propositions because the laws impose epistemic constraints on degrees of belief.

Logical truths are sentences true under all interpretations of nonlogical expressions. To make them objects of probabilities, one may take them as the propositions such sentences express. However, it is best to retain the standard, formal definition of logical truths, maintain the distinction between logical truths and a priori truths, and attend to a priori matters when applying the laws of probability. The axiom of total probability, for instance, gains scope when applied to a priori truths, which include propositions that are logical truths according to any propositional account of them. Furthermore, taking a proposition as logically true just in case a logically true sentence expresses it creates problems. In some cases, as subsequent sections show, two sentences express the same proposition although only one sentence is logically true.

Necessary truth is truth in all possible worlds. Some necessary truths, such as arithmetic truths, are not logical truths. Moreover, an identity such as "Water is H_2O" expresses a necessary truth but is not a logical truth because some interpretations give its two terms distinct denotations. An evidently true necessary proposition is an a priori truth. But not all necessary truths are evident, for instance, not the identity of water and H_2O. Because subjective probability is epistemic rather than metaphysical, a prioricity fits its axioms better than necessity does.

Probabilities yield expected utilities of an agent's options. If the agent is ideal and rational, an option's utility equals its expected utility and a decision maximizes utility. An option's expected utility is a probability-weighted average of the option's utilities relative to a set of mutually exclusive and jointly exhaustive possible states of the world, which I call a partition of states. Assuming that the states are independent of the option, the weights are the states' probabilities, and the option's utility with respect to a state is $U(o \text{ given } s)$. Therefore, an option's utility $U(o)$ equals a sum of probability-utility products, namely, $\sum_i P(s_i)U(o \text{ given } s_i)$, where i indexes the states. Weirich (2001, Chap. 3) explains this principle and the utilities it involves.

Next, I address some problems that arise in applying the laws of probability. An agent must understand the propositions to which she assigns probabilities so that she can make her assignments comply with the laws of probability. How does she attain the understanding she needs?

Imperfections in an agent's understanding of propositions may seem to generate two probabilities for the same proposition. To illustrate, reconsider the view that propositions are sets of possible worlds. According to that view, all necessary truths express the same proposition. However, given a proposition p an agent may assign probability 1 to the proposition that $p \rightarrow p$ and yet not assign the same probability to the proposition that $((p \rightarrow p) \rightarrow p) \rightarrow p$. To resolve this problem, psychological realism suggests distinguishing the two necessary truths. Taking propositions as structured, and not just sets of possible worlds, does this and provides for probability assignments less than 1 to complex necessary truths, such as the second necessary truth, which involves the material conditional more often than the first does. An agent may understand that proposition well enough to give it a probability without understanding it well enough to see that it is true. Garber (1983) revises probability theory to accommodate learning logical truths and so makes it more realistic.

This section treats another source of failure to comprehend a probability's object and advances a different response to it. To begin, consider the following case in which a rational ideal agent, who is logically omniscient, appears to assign two probabilities to the same proposition. Suppose that Frege's favorite proposition is that $7 + 5 = 12$, and imagine an agent who does not know this and assigns probability 0.5 to the truth of Frege's favorite proposition. Let the value of the propositional variable s be Frege's favorite proposition. What is the agent's probability assignment to s? His own answer depends on the proposition's presentation. He says that $P(\text{Frege's favorite proposition}) = 0.5$ and that $P(7 + 5 = 12) = 1$ and bets accordingly. It appears that he assigns two probabilities to the same proposition. His assignments may be relative to the same algebra. They may conflict only because of ignorance of empirical matters, in particular, Frege's preferences among propositions.

Besides creating an inconsistency, the assignment $P(\text{Frege's favorite proposition}) = 0.5$ violates the probability axiom that $P(s) = 1$ if s is an a priori truth. Also, it creates violations of the probability theorem that equivalent propositions have the same probability.

This problem case arises because an application of the probability laws may designate the value of a propositional variable without providing a way of grasping the proposition designated. As the example shows, some designations of a proposition do not provide access to the proposition and its probability. One way of handling the problem is to require that a probability attach to a proposition given a way of grasping it. Relativizing the probability $P(s)$ to a way of grasping the proposition s ensures access to the proposition and eliminates inconsistent probabilities.

What is a way of grasping a proposition? It is a structured way of understanding a proposition. As mentioned, my points allow some latitude in specifying its nature. It may, for instance, be a mental representation of a proposition. Standard sentential names of propositions furnish ways of grasping propositions. "The proposition that $7 + 5 = 12$" both designates a proposition and furnishes a way of grasping the proposition designated. The proposition that $7 + 5 = 12$ receives a probability assignment of 1 relative to the way of grasping it that the arithmetic formula provides. It receives no assignment relative to the designation "Frege's favorite proposition" because that designation does not furnish a way of grasping the proposition. The description does not reveal the structural features of the proposition it describes, and it designates different propositions in different worlds.

The stated grounds for relativizing probabilities to ways of grasping propositions are not compelling because there are other ways of resolving the apparent inconsistency. The remainder of this section examines an alternative response. The next section shows that it is inadequate for additional problem cases and thereby strengthens the grounds for relativization.

The alternative response to the apparent inconsistency accepts an agent's announcement of a proposition's probability only given a designation of the proposition that reveals the proposition to the agent. Hence it rejects the agent's claim that $P(\text{Frege's favorite proposition}) = 0.5$. It takes $P(\text{Frege's favorite proposition})$ as an unknown probability. The probability value the agent asserts for the proposition is the probability of a related but different proposition.

Consider (1) Frege's favorite proposition and (2) the proposition that Frege's favorite proposition is true. These propositions differ. The second proposition is not a necessary truth even if Frege's favorite proposition is that $7 + 5 = 12$. Its subject differs from world to world according to Frege's tastes, and it is false in worlds where Frege favors a false proposition. Its probability is a probability-weighted average of the worlds in which Frege favors a true proposition. Hence its probability may be 0.5 while the probability that $7 + 5 = 12$ is 1. The two propositions' probabilities also respond differently to information. $P(\text{Frege's favorite proposition is true})$, unlike $P(7 + 5 = 12)$, is sensitive to empirical information. It varies with information about Frege's tastes in propositions. "Frege's favorite proposition is true" is true just in case Frege's favorite proposition is true. However the truth predicate in the sentence quoted is not redundant; besides having a crucial grammatical role, it distinguishes the proposition that Frege's favorite proposition is true from Frege's favorite proposition. The truth predicate's role makes truth a substantial property in one sense of the phrase.

When announcing probability assignments, the agent does not reveal that he assigns to Frege's favorite proposition the probability 0.5. He reveals that he assigns that probability to the proposition that Frege's favorite proposition is true. The announced value of the unknown probability is the known probability of the metaproposition. The agent admits that he does not know what Frege's favorite proposition is. Consequently, he does not know his probability assignment to that proposition so designated. The agent does not know the value of $P(s)$ under the assignment of Frege's favorite proposition to the variable s. Its value is not 0.5 but rather 1, although the agent does not know this. He does not know his own probability assignment to Frege's favorite proposition because he does not know which proposition is Frege's favorite. The propositional name "Frege's favorite proposition" has a denotation, but the agent does not know s's value when it is assigned using that description because he does not know what proposition the description denotes. "Frege's favorite proposition" designates a proposition but does not express it and so does not offer a way of understanding it. A way of assigning a value to a propositional variable may in some cases fail to express a proposition so that the agent can grasp the proposition and can know the variable's probability given that value.

Similar cases arise if an agent runs across a proposition expressed in a foreign language. The French sentence, "Il pleut," expresses the proposition that it is raining. If the agent announces a probability assignment for the proposition the sentence expresses, he may inadvertently substitute his probability assignment to the proposition that "Il pleut" is true. The metalinguistic proposition clearly differs from the proposition that it is raining. Unlike quotations of sentences, standard sentential names of propositions rigidly designate propositions. Hence, necessarily the proposition that it is raining is true just in case it is raining. In contrast, the metalinguistic proposition is false if it is raining, given a hypothetical convention whereby the sentence "ll pleut" means that it is not raining.

For a related problem concerning utilities, consider the utility an agent assigns to a gamble g that yields \$1 if Frege's favorite proposition is true and \$0 otherwise. According to the expected utility principle, $U(g) = P(s)U(\$1) + P(\sim s)U(\$0)$ given that Frege's favorite proposition is the value of s. Using our knowledge that s is the proposition that $7 + 5 = 12$ and that the agent assigns that arithmetic proposition probability 1, we compute that $U(g) = 1U(\$1) + 0\mathrm{U}(\$0) = U(\$1)$. This is the wrong value for $U(g)$. It is not the utility assignment that should direct the agent's decision about the gamble, if, say, he has a choice between it and similar gambles involving different propositions. This application of the expected utility principle is defective because the agent does not know Frege's favorite proposition and so does not know $P(s)$ even though P is his own probability assignment.

Distinguishing (1) Frege's favorite proposition from (2) the proposition that Frege's favorite proposition is true resolves the problem. A partition consisting of the first proposition and its negation differs from a partition consisting of the second proposition and its negation. Thus, evaluating an option using the first partition differs from evaluating it using the second partition. In particular, the elements of the first partition have probabilities unknown to the agent, and their inaccessibility hampers an evaluation relying on them. A practical evaluation of a gamble concerning Frege's favorite proposition, yielding probabilities accessible to the agent, uses the second proposition and its negation. Then $U(g) = 0.5U(\$1) + 0.5U(\$0)$. Applied this way, the expected utility principle yields a utility for the gamble that appropriately guides decisions among options. Probabilities must be accessible to be factors in utilities with that guiding role.

To make relevant probabilities accessible to an agent, applications of decision principles involving probabilities typically avoid specifications of values for propositional variables that do not give the agent ways of grasping the propositions. A standard sentential expression of a proposition in the agent's language furnishes her a way of grasping the proposition. It makes the proposition's probability accessible to her and so yields a probability useful for making decisions.

The probability laws involve quantified variables. Quantifiers, under their normal interpretation, work with assignments of values to variables without discriminating among designations of those values. How can one control assignments of values to variables so that accessible probabilities result? One method adopts a substitutional interpretation of quantifiers. Then it restricts applications of the probability laws to instances that replace propositional variables with standard sentential names of propositions, as in Weirich (2001, pp. 92–100, 2004b, Sec. A.4). Another method, adopted here, retains the normal interpretation of quantifiers but requires specification of values of variables using sentential names of propositions. This method shifts the remedy from values of variables to their specifications. It uses standard sentential names of propositions to yield specifications of instances of generalizations rather than to yield the instances themselves.

Textbook applications of the probability axioms and the expected utility principle take this approach. They designate values of propositional variables so that students grasp the propositions designated. An example applying the axiom of total probability may say that the value of s is the a priori truth that Socrates died in 399 BC or Socrates did not die in 399 BC. An example applying the expected utility principle to a bet that Heads will appear on a coin toss may specify a partition of states with two members: first, the state that Heads will appear and, second, the state that Heads will not appear. Approaching applications this way dodges the problem of unknown probabilities by specifying a way of grasping each relevant proposition.

This section presented a case of apparently inconsistent probabilities and then two resolutions of the inconsistency. The first resolution relativizes probabilities to ways of grasping propositions and takes one member of the inconsistent pair of probabilities to be under-specified. The second resolution takes one probability to be unknown and to be confused with a related, known probability. To resolve apparent violations of the laws of probability, it restricts the laws to instances that furnish ways of grasping propositions. This restriction ensures access to relevant probabilities. The second resolution is not perfectly satisfactory. A variable's value is the same under any designation of that value. An assignment of a value to a propositional variable should not be limited to selected designations of propositions. Rather than press this objection, however, the next section presents another problem concerning probabilities that creates a stronger case for relativizing probabilities to ways of grasping propositions.

The Problem of Direct Reference

This section strengthens the argument for relativizing probabilities to ways of grasping propositions. In the cases it considers, a single proposition clearly receives two probabilities. Imposing restrictions on designations of values of propositional variables is not a viable alternative to relativization.

A definite description such as "the red planet" refers to an object using concepts that identify the object. Some expressions refer directly to an object without the intermediary of a concept identifying the object. For example, a variable such as "x" under an assignment of a value such as Mars refers to an object directly. The variable does not express a concept by which it refers to the object, even if the assignment of a value to the variable employs a concept. Pronouns function as variables do. On some occasions a pronoun may stand for a name introduced earlier. On other occasions, however, its denotation is not parasitic on another expression's denotation. Features of context such as pertinence and shared interests assign it a value. Consider the sentence, "It is blue," in a context where "it" designates the sky. Without expressing a concept, the subject "it" refers to an object and hence is directly referential. Frege holds that reference takes place only through the intermediary of a concept, however J. S. Mill champions the view that reference is sometimes direct, as the essays in Salmon and Soames (1988) explain.

For another example, consider demonstratives. Suppose one points to a building and says, "That's on fire." One refers directly to the building. No concept intervenes. The demonstrative has a character that defines its role in expressing a proposition, as Kaplan (1989) notes. But that character is not a concept in the traditional, Fregean sense. It is not the sort of concept a definite description expresses and is not the content of an expression. The demonstrative's content in a context is its denotation. D. Kaplan (personal communication) observes that direct reference is a means of making objects in the external world the subjects of thoughts. Direct reference enables a thought expressed in language to discriminate between the actual world and other possible worlds. A definite description may be uniquely satisfied in many possible worlds. However, the demonstrative "that" applies only to an individual in the actual world.

Proper names are not plausibly taken as abbreviations of definite descriptions; no informative definite description is synonymous with a proper name. According to the account of proper names that I adopt, they are directly referential, too. Consequently, the sentences "Cicero is Cicero" and "Tully is Cicero" have subjects directly referring to the same person. The two names may yield different ways of grasping an individual, but I do not classify those ways of grasping the individual as concepts of the individual. A concept of an individual is not just a way of picking out an individual. Concepts, according to the standard usage I follow, are abstract, shareable meanings that form propositions, and not concrete, private, psychological components of thoughts. Names such as "Sherlock Holmes" that do not denote an individual have cognitive significance but do not express concepts. The negative existential sentence, "Sherlock Holmes does not exist," does not express a proposition with Sherlock Holmes as subject.

Some predicates may yield their extensions without expressing a concept, but I assume "is blue" expresses a concept. That concept is a property according to some accounts of properties, and the objects with the property form the predicate's extension. The sentence "It is blue" expresses a proposition composed of its subject's denotation and the concept its predicate expresses. The proposition is thus a hybrid entity with a concrete component and an abstract component. The view that a proposition may contain a concrete object is Russellian. It contrasts with Frege's view that a proposition is composed of concepts exclusively. The Fregean view distinguishes the proposition that Cicero is Cicero from the proposition that Tully is Cicero, but the Russellian view of propositions that I adopt does not distinguish them. All subject–predicate sentences with subjects directly referring to the same object and having synonymous predicates express the same proposition. Differences in the way their subjects directly refer to the same object do not yield differences in the proposition they express. Hence the sentences "Cicero is Cicero" and "Tully is Cicero" express the same proposition. That proposition is a hybrid comprising a person and the concept of being identical with Cicero, which itself comprises the concept of the identity relation and Cicero. The sentences express the same proposition because they express propositions with the same components. The formula "x is Cicero" expresses that proposition under an assignment of a value to its variable. A way of grasping the variable's value completes a way of grasping the proposition.

According to Frege, the belief that Cicero is Cicero and the belief that Tully is Cicero have different objects. He holds that sentences such as "Cicero is Cicero" and "Tully is Cicero," when used to report the content of a belief, denote the propositions they normally express. Because the sentences normally express different propositions, in belief reports they denote different objects of belief. The theory of direct reference challenges Frege's analysis of propositions and belief. It holds that "Cicero is Cicero" and "Tully is Cicero" express the same proposition and so a single object of belief.

Two sentences differing only in means of directly referring to their subjects offer two ways of grasping the same proposition. Do two ways of understanding the same proposition emerge in other cases also? Are there multiple ways of expressing the same proposition without direct reference? "Female foxes are female foxes" and "Vixens are female foxes" come close to being synonymous. Perhaps they express different propositions, however, because it appears that a person may believe one and doubt the other. The appearance may be an illusion. If a person affirms the first sentence and denies the second, perhaps he does not understand both sentences. Maybe he denies the second sentence because he does not understand the word "vixen." His affirming and denying sentences does not establish his attitude toward propositions they express unless he understands the sentences.

A. Church (personal communication) holds that "$\forall x \mathrm{F}x$" is synonymous with "$\forall y \mathrm{F}y$." If one sincerely asserts the first and denies the second, then one does not understand both. If one understands both, then to believe one is to believe the other. Perhaps such formal synonymies do not yield two ways of understanding the same proposition. They may offer the same way of understanding a proposition because they lead to the same propositional attitudes and never to believing the proposition understood according its first expression and disbelieving the proposition understood according to its second expression.

To believe a proposition or assign it a probability, an agent must understand the proposition. An agent may understand it without having perfect knowledge of the concepts it involves. Burge (1979) observes, for example, that one may possess the concept of arthritis without understanding it perfectly and, in particular, may have beliefs about arthritis without knowing that it occurs only in joints. An agent may, for instance, believe that she has arthritis in her thigh. Medical concepts are introduced by theories that employ them. One can grasp a concept without knowing the entire theory that introduces it. Arthritis has an implicit definition via medicine, but one may form beliefs involving the concept without being an expert in medicine. Perhaps imperfect understanding of the concept arthritis may lead one to believe and doubt a proposition involving that concept. The two attitudes may draw on different parts of one's understanding of the concept, each of which is sufficient for understanding the proposition. Perhaps through imperfect understanding an agent may both believe and doubt that she has arthritis.

Although direct reference may not be the only source of multiple ways of understanding the same proposition, I rely only on it in making a case for relativizing probabilities. It yields clear cases of propositions understood two ways.

The directly referential sentences "I have mustard on my chin" and "He has mustard on his chin" express the same proposition if I utter the second while pointing to my image in a mirror. These two ways of expressing (and not simply designating) the same proposition supply two ways of understanding that proposition. Hence I may believe and disbelieve the proposition, or assign it different probabilities, according to my way of grasping it. If I do not know that the image in the mirror is mine, I may disbelieve the proposition when I entertain it using the expression "I have mustard on my chin" but believe it when I entertain it using the expression "He has mustard on his chin."

Tradition treats propositions as basic bearers of truth-values and also objects of thought. Can propositions serve both roles? The propositions that Cicero is Cicero and that Tully is Cicero have the same truth-conditions and are identical. Yet an agent may assert that Cicero is Cicero and deny that Tully is Cicero. His attitudes appear to have different objects. If the object of a belief is a proposition, how can this happen? Can an agent believe and disbelieve the same proposition? Making propositions both objects of thought and basic bearers of truth-values gives propositions two roles that they cannot serve equally well. Propositions cannot be attuned to both cognitive values and truth-conditions because cognitive values and truth-conditions are not attuned to each other.

What role does traditional usage favor? In my view, the primary role of propositions is to be basic bearers of truth-values. Instead of distinguishing two types of proposition, one for each role, I follow a usage that makes the primary role definitive. In the example concerning Cicero, there is just one proposition but two ways of expressing it with different cognitive values.

Disbelief is not failure to believe, that is, the absence of an attitude, but rather doubt, an attitude toward a proposition. In the example concerning Cicero, the agent does not believe and fail to believe the same proposition. That is impossible. Instead, the agent believes and disbelieves the same proposition. The agent, not the example, is inconsistent. If, contrary to my analysis, disbelief is treated as failure to believe, then the example may be recast so that it involves believing both a proposition and its negation. According to the theory of direct reference, the proposition that Tully is not Cicero is the negation of the proposition that Cicero is Cicero. An agent may believe that Tully is not Cicero. He may understand the proposition but not see that it is false. Then the agent may believe that Cicero is Cicero and also believe that Tully is not Cicero. No contradiction arises from an agent's believing both a proposition and its negation. An agent's holding contradictory beliefs does not create a contradiction.

In my example, I assume that the agent has in his vocabulary the two names "Cicero" and "Tully." Perhaps he is a beginning student of the classics who has come across both names but has not yet realized that they are names for the same person. Understanding a proposition requires knowing the proposition's meaning but does not require knowing everything relevant to the proposition's truth. An agent may understand the proposition that Tully is Cicero without knowing that Tully is Cicero. He may know the person "Tully" designates, being familiar with the name, but may not know everything about that person, in particular, that "Cicero" denotes that person. He may believe that Cicero is Cicero and yet disbelieve that Tully is Cicero. Then he believes and disbelieves the same proposition. He may understand it well enough to believe it, and doubt it, without knowing everything about the proposition, in particular, that he grasps it two ways.

A good ground for attributing a belief to an agent is the agent's sincere assent to a sentence of his language expressing the proposition. Similarly, sincere denial is a good ground for attributing disbelief. To fill out the example, let the agent described assent to the sentence "Cicero is Cicero" and deny the sentence "Tully is Cicero." Then he assents to a sentence expressing a proposition and denies another sentence expressing the same proposition. This all happens without a change in circumstance grounding a change in opinion. Although the two sentences express the same proposition, they have different cognitive significance. The way of grasping the proposition the first sentence furnishes makes the proposition's truth apparent. That explains why an agent may affirm the first sentence and yet deny the second sentence.

A report of an agent's belief may imperfectly indicate the belief's content. A journalist may use English to report in translation a Russian official's speech. The report, although true, may only imperfectly indicate the propositions the official believes. Also, a true report of a belief may use a brief description of the proposition believed. It may say, "He believes what Bush said," or "He believes that theory's first proposition." To be true, a report of a belief must present the proposition believed in a way that identifies the proposition but not necessarily in a way that permits recipients of the report to grasp the proposition. Perhaps the report that the agent believes that Tully is Cicero is true, because he believes that Cicero is Cicero, but is misleading because he denies that Tully is Cicero.

The distinction between beliefs and reports of beliefs is important but not adequate for resolving the problem of multiple ways of grasping the same proposition. To explain, let me compare an agent's not knowing that the sentences "Cicero is Cicero" and "Tully is Cicero" express the same proposition with other cases of ignorance concerning propositions. In the first case an agent does not know that the sentence "It is raining" and the French sentence "Il pleut" express the same proposition. In the second he does not know that the descriptions "the proposition that $7 + 5 = 12$" and "Frege's favorite proposition" name the same proposition. In the third he does not know that the statements "He was a genius" and "Einstein was a genius" express the same proposition because he does not know that the first statement refers to Einstein in its context. In these cases the agent does not believe a proposition given one way of grasping it and disbelieve it given another way of grasping it. Rather, he does not know what proposition a linguistic item presents. As the section "Applications of Probability Laws" explains, in the second case the agent does not know the denotation of "Frege's favorite proposition" and so does not grasp any proposition by means of that designation. If he says that Frege's favorite proposition is false, he is not expressing disbelief that $7 + 5 = 12$. Similarly, in the other two cases, a failure to endorse a proposition believed, when presented a certain way, does not indicate disbelief but rather ignorance of the proposition presented.

Direct reference creates belief and disbelief concerning the same proposition. The sentences "Cicero is Cicero" and "Tully is Cicero" furnish two ways of grasping the same proposition. They are full expressions of a proposition and are in the agent's repertoire. In contrast, the description "Frege's favorite proposition" does not provide the agent a way of grasping the proposition that $7+5 = 12$. The French sentence "Il pleut" is not in the agent's repertoire. The statement "He was a genius" does not furnish the agent a way of grasping a proposition because he is not aware of the features of its context that yield the subject's denotation. Because the sentences "Cicero is Cicero" and "Tully is Cicero" furnish two ways of grasping the same proposition, it is implausible that the agent withholds assent to the latter because he is ignorant of the proposition it expresses. It is implausible that he believes that Tully is Cicero but does not know he holds that belief. It is implausible that a report of that belief is just misleading and not also false.

Direct reference creates complications for degrees of belief as well as for belief. Because the proposition that Tully is Cicero is the same as the proposition that Cicero is Cicero, it seems that P(Tully is Cicero) equals P(Cicero is Cicero). But an agent may not know that the propositions are the same. Because their identity depends on an empirical matter, even a rational ideal agent may be ignorant of it. Such an agent may be certain that Cicero is Cicero and yet doubt that Tully is Cicero.

Sincere assertion of the probability of a proposition a sentence expresses is good evidence that the proposition's probability has the value asserted, provided that the sentence is in the agent's language. If when asked about his probability assignment for the proposition expressed by "Cicero is Cicero," a sincere agent says 1, and if when asked about his assignment for the proposition expressed by "Tully is Cicero," he says 0.5, then it is plausible that the probabilities announced are the probabilities the agent assigns. In such cases it appears that the agent assigns two probabilities to the same proposition and so is inconsistent. The two assignments are psychologically compatible because each assignment represents a degree of belief and not the absence of other degrees of belief.

The problem of direct reference runs deeper than the problem of unknown probabilities exemplified by the case of P(Frege's favorite proposition). Designating values for propositional variables in ways that reveal propositions does not resolve it. The sentence "Tully is Cicero" furnishes a way of grasping a proposition. So P(Tully is Cicero) is not an unknown probability. One cannot discard or reinterpret it, as the section on Applications of Probability Laws handled P(Frege's favorite proposition). In contrast with apparent inconsistencies arising from ignorance of propositions, the appearance of two probabilities for the same proposition in cases of direct reference cannot be dispelled by requiring applications to assign a value to a propositional variable using a standard name of a proposition. Direct reference creates two standard names of the same proposition. The sentences "Cicero is Cicero" and "Tully is Cicero" are equally good expressions of a proposition although they differ in cognitive significance. The problem of direct reference remains despite requiring that applications of probability laws fix values of propositional variables using ways the agent grasps propositions. The requirement discards opaque designations of propositions, such as "Frege's favorite proposition," because they make probabilities inaccessible. However, direct reference creates distinct full expressions of a single proposition. Assigning to a propositional variable the proposition that Tully is Cicero is just as transparent as assigning the proposition that Cicero is Cicero. Both sentential expressions of the proposition are standard names of propositions. Both furnish ways of grasping a proposition. Both are in the agent's repertoire. There are no grounds for dismissing the assignment of the proposition that Tully is Cicero as there are for dismissing the assignment of Frege's favorite proposition. There are no grounds for saying that the agent does not identify the proposition that Tully is Cicero. He identifies the proposition even if he does not know that it is the same as the proposition that Cicero is Cicero. The problem of two probabilities for the same proposition, when it arises from direct reference, cannot be resolved by declaring one probability to be unknown.

Direct reference creates cases in which an agent knows a proposition and its probability given a way of grasping it but does not know that two ways of grasping a proposition yield the same proposition. Using independently the two ways of grasping the proposition, he assigns it two probabilities. In the example, the agent knows the proposition that Cicero is Cicero and knows the proposition that Tully is Cicero but does not know that they are the same proposition because he does not know that Tully and Cicero are the same person.

Suppose that because of direct reference a proposition has two probabilities, each according to a different way of grasping the proposition. That plainly violates the laws of probability. Inconsistent probabilities also upset principles of rational choice. The expected utility principle licenses use of any partition of states. To instantiate the principle, a state variable s_i is assigned a proposition as value. The assignment is indifferent among ways of grasping the proposition. However, to use the principle's instance, an agent adopts a way of grasping the proposition assigned.

One has not specified a propositional variable's value adequately for applications unless one specifies a way for the agent to grasp the proposition. That ensures the agent's access to its probability but also makes the principle's application sensitive to various ways of grasping the same proposition.

Suppose that an agent frames a decision problem using certain ways of grasping propositions that form a partition of possible states. Those ways of grasping the propositions influence his assignment of probabilities to the propositions. Then the probabilities influence the expected utilities of options and whether an option maximizes expected utility. An inconsistency in decision recommendations arises if the agent also frames the decision problem using different ways of grasping the same states, and they generate different probabilities for the states and therefore a different ranking of options according to expected utilities.

To resolve the problem of direct reference, one must deal with the problems it causes for the laws of probability and the principle of expected utility maximization. Later sections make a proposition's probability relative to a way of grasping the proposition. In my view, the inconsistencies in probabilities that direct reference creates are not defects in assignments of probabilities but defects in representations of probabilities.

Alternatives to Relativization

The laws of probability applied to degrees of belief appeal to idealizations. For instance, they treat the degrees of belief of ideal agents, who are logically omniscient. Should the laws incorporate the additional idealization that agents know for any two ways of grasping a single proposition that they yield the same proposition? Because such knowledge requires empirical information, demanding it is counter-productive. The laws' objective is to treat agents with incomplete empirical information, and so they must accommodate ignorance that two ways of grasping a proposition yield the same proposition. They cannot meet their objective if they dismiss such ignorance. Similarly, restricting the laws to propositions whose probabilities are constant with respect to ways of grasping them is counter-productive. To furnish a general treatment of degrees of belief, the laws must dispense with that restriction.

I seek a resolution of the problem of direct reference that minimally revises probability theory. Relativizing a proposition's probability to a way of grasping the proposition does this. It allows an agent's grasp of a proposition to influence the probability she assigns the proposition, just as her evidence influences that assignment. It gives her access to her assignment and allows her to state it when furnished the way of grasping the proposition.

A formally equivalent alternative to this relativization takes a probability's object to be a pair composed of a proposition and a way of grasping it. If a probability's object were expanded this way, then probability's attention to truth would narrow to the truth of the pair's first component. The second component would just furnish a means of assessing the first component's truth.

Although a conservative response to the problem of direct reference may take various forms, some responses to the problem are not conservative. They necessitate major changes in probability theory. This section reviews their drawbacks.

One response to the problem of direct reference is to make the objects of probabilities finer-grained than propositions. Let them be ways of grasping propositions, for instance. “Cicero is Cicero” and “Tully is Cicero” yield two ways of grasping the same proposition. Let those ways of grasping the proposition be objects of probabilities. Their probabilities may be the probabilities that they yield true propositions. The literature suggests many ways of fleshing out this response. A way of grasping a proposition might be a narrow rather than a wide content of a thought with internal rather than external individuation. It might be a concrete, particular mental representation, and might involve private rather than socially shared concepts.

Suppose that ways of grasping propositions are sets of centered-worlds. Lewis (1983) defines a centered world as a pair consisting of a world and an individual in the world. Chalmers (2002) adds a time in the world to the world’s center. Other definitions are possible, too. Specifying a world’s center accommodates direct reference to individuals. However, sets of centered worlds do not have enough structure to distinguish necessary truths. Moreover, the sets do not distinguish the ways of grasping a proposition that the sentences “He is in danger” and “I am in danger” furnish in cases where their subjects refer to the same individual.

Suppose that ways of grasping propositions are epistemic intensions, as Chalmers (2002) characterizes them. These are functions from scenarios, or epistemic possibilities, to truth-values, and so may be represented by sets of scenarios. Scenarios are representations of centered worlds. Hence the objections to taking sets of centered worlds as objects of probability carry over to epistemic intensions. Also, epistemic intensions are a type of narrow content. Truth according to a scenario is not truth according to a centered possible world, for a scenario may contain impossibilities such as the non-identity of Tully and Cicero. Not being bearers of truth-values in the ordinary sense, epistemic intensions are not apt for probability assignments. Chalmers, in fact, takes belief to attach not to epistemic intensions but rather to two-dimensional intensions that comprise both epistemic intensions and standard intensions going from possible worlds to truth-values. His two-dimensional account of belief resembles a relativization of propositional belief to ways of grasping propositions.

Several general problems also caution against taking ways of grasping propositions as objects of probabilities. These problems do not presume any particular interpretation of ways of grasping propositions. First, it may turn out that a single way of grasping a proposition yields different propositions in different contexts. This happens if factors such as the character of an indexical are involved in ways of grasping propositions. Then the probability assigned to the way of grasping a proposition must be relative to a context. However, applying the probability laws requires combining objects of probability. How can one combine ways of grasping propositions relative to different contexts? For example, suppose that “I am hungry” furnishes a way of grasping a proposition, and it is relative to contexts including times. A combination of its uses at different times is not relative to any time, but needs relativization to yield an object of probability. Instead of devising a way of combining context-relative ways of grasping propositions, it is simpler to combine propositions, as traditional probability theory does. It is more conservative to assign probabilities to propositions relative to ways of grasping them.

Second, it may turn out that ways of grasping propositions are private, not shared among people. Then making them objects of probabilities makes it hard to explain cases in which we say two people assign a probability to the same object. Cases of agreement and disagreement concerning probabilities become complicated. It is simpler to treat them as cases in which people appraise the same proposition, each according to his grasp of it.

Third, in special cases an agent may assign a probability to a proposition without adopting a way of grasping it. For example, an agent may assign the same probability to every atomic proposition of an algebra, thus distributing probability uniformly over the atoms, without entertaining each proposition individually. This may happen if propositions are objects of probability under a null relativization to ways of grasping propositions. But it cannot happen if ways of grasping propositions are objects of probability.

Fourth, taking ways of grasping propositions as objects of probability disregards probability's focus on truth. Ways of grasping propositions are not fine-grained propositions. Even if represented by functions from worlds to truth-values, they yield a truth-value only by first yielding a proposition. Ways of grasping a proposition, if assigned probabilities, have probabilities parasitic on the probabilities of propositions. Therefore, nothing grounds two probability assignments to two ways of grasping the same proposition. Even taking account of structural considerations, the same truth-conditions govern the propositions provided by the sentences "Cicero is Cicero" and "Tully is Cicero." Associated ways of grasping propositions have the same derivative truth-values because those truth-values derive from the same proposition. Even if the two ways of grasping a proposition are distinct objects of probability, their probabilities' attention to the same truth-conditions align their values. The cognitive difference between the expressions of propositions does not yield a difference in truth-conditions to ground a difference in probability assignment. The objects of probabilities cannot differ in probability despite being governed by the same truth-conditions unless their probabilities are relative to different contextual features. Simply making the objects of probability fine-grained, without relativizing probabilities to fine-grained contextual features, leaves the problem of direct reference unresolved.

Another rival of relativization, taking the objects of probabilities to be sentences interpreted by contexts, faces a similar objection. The probabilities of sentences are parasitic on the probabilities of propositions they express. Suppose that two sentences express the same proposition although an agent does not realize this. For instance, an agent may not know that "It is raining" and the French sentence "Il pleut" express the same proposition. Nothing grounds an assignment of different probabilities to those sentences. Because they express the same proposition, they have the same truth-conditions. Probability assignments attend only to those truth-conditions and generate the same value for each sentence.

In general, although thoughts and sentences bear truth-values, propositions are the basic bearers of truth-values. The truth-values of propositions yield the truth-values of sentences and thoughts. So making thoughts or sentences the objects of probabilities does not handle cases in which an agent assigns two probabilities to the same proposition. This phenomenon demands another revision. The revision must change the interpretation of probability to make a proposition's probability depend not only on the proposition but also on a way of grasping it.

To conclude this section, I examine an attempt to resolve the problem of direct reference by appealing to metalinguistic propositions. This approach takes belief that Cicero is Cicero and doubt that Tully is Cicero as, respectively, belief that "Cicero is Cicero" is true and doubt that "Tully is Cicero" is true. According to it, objects of probabilities are not sentences but propositions about sentences. Although the truth-conditions for the two sentences mentioned are the same, the truth-conditions for the propositions about the two sentences differ. For instance, the propositions' truth-values depend on different empirical facts about the denotations of names.

This metalinguistic approach fails to resolve the problem of direct reference. First, it is unfaithful to the phenomenon of belief. Suppose that "*p*" abbreviates an English sentence (and to dispense with corner quotations marks suppose that the abbreviation persists within ordinary quotation marks). Belief that *p* is not the same as belief that "*p*" is true. That expansion generates an endless ascent of metabeliefs. According to it, belief that "*p*" is true is belief that "'*p*' is true" is true, and so on. Second, the objections to taking belief's objects to be sentences rather than propositions carry over to the view that belief's objects are propositions about sentences being true. For example, for a French speaker, belief that it is raining is not belief that "It is raining" is true. Third, the problem of direct reference resurfaces at the metalevel in cases with two names for the same sentence. A person may believe that "$7 + 5 = 12$" is true and disbelieve that Frege's favorite sentence is true although "$7 + 5 = 12$" is Frege's favorite sentence. Similarly, an agent may affirm the sentence "Cicero is Cicero" and deny the sentence "That sentence is true" in a context where "that sentence" denotes "Cicero is Cicero." He may not know that the demonstrative has that denotation. For these reasons, probability theory needs another approach to the problem of direct reference.

Relativization

All accounts of subjective probability recognize that an assignment of a probability to a proposition is relative to an agent and to a time. To resolve the problem of direct reference, I make it relative also to a way of grasping the proposition. Accordingly, an agent assigns a probability to a proposition at a time given a way of grasping the proposition. This relativization of probability assignments still takes the object of a probability to be a proposition but lets a contextual factor influence that assignment.

Crimmins (1992) takes belief as a relation between an agent, time, proposition, and mental representation of the proposition. According to Crimmins, a report that

Caius believes that Cicero is Cicero may be true while a report that Caius believes that Tully is Cicero is false because the reports supply different mental representations to which the beliefs reported are relative. In particular, the report that Caius believes that Tully is Cicero may be a false report because the context it creates indicates a mental representation by which Caius does not belief the proposition. Weirich (2004a) extends Crimmins' account of belief to degrees of belief. This section elaborates the extension's consequences for the laws of probability. However, it replaces a mental representation of a proposition with a way of grasping the proposition. A way of grasping a proposition may be interpreted, following Crimmins, as a concrete particular mental representation but may be interpreted other ways, too, for example, as abstract and shareable. To maintain neutrality about side issues, I do not articulate an account of a way of grasping a proposition.

McKinsey (1999) uses "mode of presentation of a proposition," a term roughly interchangeable with "way of grasping a proposition," to mean a traditional Fregean sense. However, I use the technical term "way of grasping a proposition" without any commitment to Frege's view. Using the term does not indicate a stand on issues distinguishing a mode of presentation of a proposition, a guise in which a proposition appears, and a mental representation of a proposition.

My introduction of ways of grasping a proposition is minimal but clarifies some important points. A way of grasping a proposition enables an agent to understand a proposition, to know it, and to identify it. A sentential expression of a proposition provides a standard way of grasping a proposition. Also, I assume that ways of grasping propositions may be combined to form compound ways of grasping compound propositions. I use logical operations such as disjunction to form such compounds.

Let $P(s,\ w)$ stand for the probability of a proposition s given a way w of grasping that proposition. Consider an agent's probability assignment to the proposition that Cicero is Cicero and to the proposition that Tully is Cicero in a case where the agent is certain of the proposition understood using its first expression but doubts the proposition understood using its second expression. In this case $P(C = C,\ w_1) \neq P(T = C,\ w_2)$. The probabilities concern a single proposition, but the ways of grasping that proposition differ. That difference explains the difference in probabilities.

How does making probabilities relative to ways of grasping propositions affect the laws of probability? First, it prevents the embarrassment of assigning two probabilities to the same proposition relative to the same indices. It allows the laws to maintain single-valued probability assignments. Second, it requires rewriting the laws of probability to accommodate the extra index. The expected utility principle also needs rewriting to accommodate relativizing a state's probability to a way of grasping the state. The resultant relativization may generate different rankings of options relative to different ways of framing states.

The standard probability laws handle the indices of person and time by fixing them. The probability laws apply to just one person's degrees of belief at just one time. Can the index for a way of grasping a proposition be fixed, too? One might specify for each proposition exactly one way of grasping the proposition, perhaps a way of grasping it indicated by a standard sentential expression of the proposition. Textbook problems do this implicitly. However, this approach accommodates relativization by making it irrelevant. I want to preserve the relevance of relativization and use it to address the problem of direct reference. I seek an approach that handles cases in which the same proposition receives different probabilities given different ways of grasping it. Such cases require rewriting the laws of probability and the principle of expected utility maximization.

Applications of the probability law that $P(s) = 1$ if s is an a priori truth may not supply, along with a proposition assigned as a value for the variable s, a way of grasping the proposition. I supply the missing factor by adding to the law another variable that takes as a value a way of grasping a proposition. The additional variable makes the law more sensitive: $P(s, w) = 1$ if s is an a priori truth when grasped in way w. The relativization also governs the a priori status of a proposition. The proposition s is a priori relative to a way of grasping it. It is a priori relative to w if grasped via w its truth-value is apparent to any rational ideal agent. It does not matter whether employing w requires having had certain experiences.

The standard probability laws presume that a probability assignment is a single-valued function that yields just one probability for a proposition. The laws do not regulate a set-valued or multi-valued function and for such functions leave unspecified relations between probabilities such as $P(s)$ and $P(\sim s)$. Independently relativizing each proposition's probability to a way of grasping it weakens the laws' inferential power. The same holds for the expected utility principle's inferential power. Relativizing for multiple ways of grasping the same proposition risks destroying the structure standard principles impose on probabilities and decisions. Relativizing should take care to preserve that structure as much as possible.

Sometimes, according to an agent's epistemic standards, his evidence is insufficient for precise probability assignments. For instance, the probability of rain may have admissible values ranging from 0.3 to 0.7. Skyrms (1990, pp. 66–67) uses sets of entire probability assignments to represent imprecise probability judgments. Each admissible value of a proposition's probability is relative to an admissible total probability assignment. This representation of imprecise probability judgments preserves the structure the probability laws impose on probabilities relative to the same total probability assignment. Relativization to a total probability assignment accommodates imprecision without losing traditional structure.

To preserve structure among probabilities, relativization for intensionality must also be coordinated. It cannot yield probabilities that are independently relative to ways of grasping propositions. Preserving structure requires relativizing complete probability assignments to ways of grasping propositions. A proposition receives several probabilities with respect to several ways of grasping it only in the context of several total probability assignments, each of which obeys the standard laws. As a result, given each total probability assignment $P(s, w)$ equals $1 - P(\sim s, \sim w)$, where $\sim w$ is a compound way of grasping $\sim s$ formed from w using negation. The other laws relativized for ways of grasping propositions similarly hold with respect to total probability assignments.

The next section rewrites the probability laws relativizing a proposition's probability first to ways of grasping a proposition and then to total probability assignments. The relativization to total probability assignments rebuilds structure after relativization to ways of grasping a proposition breaks it down.

Relativizing the Laws of Probability

To set the stage for rewriting the probability laws, let me note how the phenomenon of direct reference bears on their basic design, in particular, on their application to propositions and their attention to a priori features of propositions.

In applications of the probability laws to sentences, an algebra generated by atomic sentences yields the compound sentences to which probabilities attach. Suppose that "C" stands for Cicero and "T" stands for Tully. If an atomic sentence s is "$C = C$," then $\sim s$ is "$C \neq C$" and not "$T \neq C$." Hence if the probability laws treat sentences instead of propositions, they do not need rewriting to handle multiple ways of grasping a proposition. Applying the laws to sentences in effect restricts them to a single way of grasping each proposition. Retreating to assignments of probabilities to sentences thus dodges the problem of direct reference. Nonetheless, that retreat is unappealing. The laws gain generality when applied to propositions, and that gain compensates for wrestling with direct reference.

Also, treating logical truth as a classification for propositions, if "$C = C$" expresses a logical truth so does "$T = C$" because the sentences express the same proposition. Granting that logical truths receive maximum probability, the proposition each sentence expresses has a probability of 1. This is an awkward consequence of treating logical truths as propositions. The phenomenon of direct reference therefore counts against that view. It reinforces the reasons in the section on Applications of Probability Laws for taking logical truths as sentences and making the probability laws attend to a priori features of propositions.

Because the same proposition may be grasped in ways differing in cognitive significance, a proposition's classification as an a priori truth is relative to a way of grasping the proposition. Take the proposition that Cicero is Cicero. By itself, it is not classified as a priori or not a priori. Its status is relative to a way of grasping it. Although it is a necessary truth independently of ways of grasping it, its epistemic status is relative to a way of grasping it typically furnished by its expression. The proposition that Cicero is Cicero has a priori status grasped as a self-identity but not grasped using "T = C." It may not be a priori even grasped using "C = C" if the agent is unaware of the co-referentiality of the two occurrences of "C" in that sentence. The a priori inconsistency of a pair of propositions is similarly relative to ways of grasping its elements. The propositions that Cicero is Cicero and that Cicero is not Cicero are not inconsistent a priori relative to ways of grasping those propositions using "C = C" and "T ≠ C." Relative to grasping the pair as a proposition and its denial, however, they are inconsistent a priori. Given those ways of grasping the propositions, their inconsistency is apparent to a rational ideal agent.

To simplify relativization of the probability laws, I reformulate them for cognitively unlimited agents only and postpone their extension to cognitively limited agents. Also, I assume that every relevant proposition has a precise probability assignment. Standard methods of accommodating imprecision may add generality later. I adopt a notational convention, too. If w is a way of grasping s and v is a way of grasping r, then $w \vee v$ is a way of grasping $s \vee r$.

Given these stipulations, making probabilities relative to ways of grasping propositions yields the following new axioms:

For all s and w such that w is a way of grasping s, $P(s,\ w) \geq 0$.
For all s and w such that s is an a priori truth grasped via w, $P(s,\ w) = 1$.
For all s, w, r, v such that s grasped via w and r grasped via v are inconsistent a priori, $P(s \vee r,\ w \vee v) = P(s,\ w) + P(r,\ v)$.

The theorem stated earlier becomes: For all s, w, r, v such that s grasped via w and r grasped via v are equivalent a priori, $P(s,\ w) = P(r,\ v)$.

As the previous section mentioned, this relativization to ways of grasping propositions puts in jeopardy proofs of traditional theorems because the proofs use a priori properties and relations of propositions. For example, the proof of the theorem that $P(\sim s) = 1 - P(s)$ uses the a priori truth of $s \vee \sim s$ and the a priori inconsistency of the disjuncts. Perhaps $s \vee \sim s$ is not an a priori truth if s is grasped via w and $\sim s$ is grasped via $\sim v$, where v is a way of grasping s different from w. To deal with the problem, one may fix the way of grasping s throughout the theorem's proof. Let the theorem say that $P(\sim s,\ \sim w) = 1 - P(s,\ w)$ where $\sim w$ is a way of grasping $\sim s$ via negation and w. Then appealing to omniscience about a priori matters, the proof succeeds. For an agent with that omniscience, $s \vee \sim s$ is an a priori truth if grasped via $w \vee \sim w$. Similarly, s and $\sim s$ are inconsistent a priori if grasped via w and $\sim w$ respectively.

How may one fix ways of grasping propositions so that proofs of theorems succeed? Traditional laws make a proposition's probability relative to an algebra containing it. Let a way of grasping each atomic proposition accompany the proposition in the algebra that furnishes the objects of probability. Then relativize each compound proposition of the algebra to a way of grasping it obtained from analogous compounding of the ways of grasping the atomic propositions that generate it. The compound way of grasping the compound proposition mirrors the structure of the proposition. Relativized this way, the new laws of probability govern a rational agent's total assignment of probabilities relative to an algebra. They fix a way of grasping a proposition throughout inferences involving the probability laws.

This relativization accommodates the problem of direct reference without sacrificing the structure the standard probability laws impose. It just generalizes the standard probability laws. The standard laws are recoverable as a special case of the new laws by fixing the ways in which atomic propositions are grasped. Relativizing to ways of grasping propositions does not destroy the structure the probability laws impose. It just elaborates the laws' traditional relativization to an algebra of propositions.

How much freedom does the relativization allow? Whereas the standard presentation of the probability laws presumes that a rational ideal agent's assignment of probability to a proposition will not vary from algebra to algebra, I allow the assignment to vary if acceptable ignorance of empirical matters accounts for the variation.

To illustrate the new probability laws, let me apply them to the example concerning beliefs about Cicero and Tully. To begin, consider the law concerning propositions that are true a priori grasped a certain way. Consider the proposition that Cicero is Cicero. Grasped as asserting a self-identity, it is an a priori truth, and so has probability 1. In contrast, grasped using "Tully is Cicero," it is not an a priori truth and so may have probability less than 1. Also, the proposition grasped as a self-identity is not a priori equivalent to the proposition grasped using "Tully is Cicero." So the proposition may receive different probabilities under those two ways of grasping it without violating the theorem concerning a priori equivalence. To use the probability laws to form degrees of belief, an agent must know how she grasps propositions. The cognitively ideal agents the laws govern have this knowledge. However, even an ideal agent may not know that she grasps the same proposition two ways.

Although I have relativized probability to an algebra of propositions grasped certain ways, I have not relativized probability to ways of grasping the truth-functions generating the algebra. This is not necessary because an ideal agent grasps a truth-function only if she grasps its effect on the truth-conditions of compound propositions formed with it. Consequently, a rational ideal agent does not assign different probabilities to a compound proposition according to different ways of grasping the truth-functions it involves.

Expected Utility

Kahneman and Tversky (1979, pp. 286–288, 1982, pp. 166–168) demonstrate the influence on decisions of frames for decision problems. Take, for example, the case of the rare Asian disease. The disease strikes 600 people. You are a public health official and must decide on a course of treatment.

Situation I. The available treatments have the following outcomes:
Treatment A. 200 people are saved for sure.
Treatment B. A probability 1/3 of saving all 600 and a probability 2/3 of saving 0.
Situation II. The available treatments have the following outcomes:
Treatment C. 400 people die for sure.
Treatment D. A probability 1/3 that none die and a probability 2/3 that all 600 die.

Most people pick treatment A over treatment B in situation I, and treatment D over treatment C in situation II. But the situations are the same and are only described differently. Treatment A is equivalent to treatment C, and treatment B is equivalent to treatment D. Situation II represents outcomes as losses, whereas Situation I represents them as gains with respect to a lower reference point. The two ways of framing the same decision problem create inconsistent preferences. People are more ready to take chances to avert possible losses than to secure possible gains. So representing outcomes as losses rather than gains increases willingness to take risks.

Many other decision theorists also observe that different ways of framing decision problems lead to different choices, for example, Resnik (1987, pp. 8–10), Rubinstein (1998, p. 17), van Fraassen (2002, pp. 103–108), Peterson (2003, pp. 23–33), and Schick (2003, Chap. 2). In cases where unequally valued choices emerge from two representations of the same decision problem, usually at least one representation, or its employment, is mistaken. Rationality does not support both decisions. In particular, the framing effects Kahneman and Tversky describe disappear if the agent recognizes the logical equivalence of different ways of framing the same decision problem. Logic and a priori knowledge eliminate common framing effects.

Although for an ideal agent with precise probability and utility assignments a decision's rationality is generally independent of framing, direct reference creates exceptional cases. It yields multiple ways of understanding the same proposition. A rational agent may understand a proposition two ways without realizing that each way yields the same proposition. The probability an agent assigns to a state depends on his way of grasping it. Each way of grasping it provides an understanding of the proposition, even if some ways of grasping it better display its truth-value. Rational framing effects are possible, therefore. The framing effects direct reference generates do not disappear given logical omniscience and perfect cognitive capacities. They arise because of ignorance of empirical matters.

In preferences concerning treatments of the Asian disease, a typical agent understands the treatments and their possible outcomes but does not recognize that some outcomes are logically equivalent. Limited logical perspicuity may excuse that failure. Other decision problems generate another type of excuse. Suppose that an agent can buy for \$0.50 a gamble that pays \$1.00 if Frege's favorite proposition is true and nothing otherwise. If he does not know Frege's favorite proposition, he decides without understanding the proposition on which the outcome turns. That ignorance may excuse a failure to buy the gamble although it guarantees a profit of \$0.50 because Frege's favorite proposition is that $7 + 5 = 12$.

Cases involving direct reference put aside both types of excuses. Suppose an agent is ideal and understands the propositions at issue. He declines an opportunity to buy for \$0.50 a gamble that pays \$1.00 if Tully is Cicero and nothing otherwise. Although he understands the bet and reasons flawlessly, he is ignorant of relevant empirical matters. He does not know that Tully is Cicero, and so passes up a chance to guarantee \$0.50 profit. He assigns a probability of 0.5 to the proposition at issue, grasping it as a general identity rather than as a self-identity.

The decision's rationality depends on the way the agent frames the decision problem. His decision has a proposition as object, a first-person action proposition. The agent may grasp the proposition various ways. His decision's rationality depends not only on the proposition but his grasp of it. The proposition that I bet that Cicero is Cicero is identical to the proposition that I bet that Tully is Cicero. Nonetheless, the presentations of the proposition have different cognitive significance. Declining the gamble is irrational if the gamble is framed as a wager that Cicero is Cicero but is rational if the gamble is framed as a wager that Tully is Cicero. Indeed, the gamble has different expected utilities according to different ways of grasping the states that generate its expected utility. For example, one partition of states presents the state that Cicero = Cicero and the state that Cicero $\neq$ Cicero. Another partition presents the state that Tully = Cicero and the state that Tully $\neq$ Cicero. The expected utilities of options vary with the partition presented when probabilities of states are relativized to ways of grasping states. It may be rational to bet that Cicero = Cicero and not to bet that Tully = Cicero because of framing effects. That is, one may rationally accept and reject the same bet offered twice if the offers present the bet differently.

This example assumes that bets are individuated according to propositions bet upon without regard for the way in which the propositions are grasped. A bet that Cicero = Cicero is the same as a bet that Tully = Cicero despite differences in the pivotal proposition's presentation. The proposition's truth-value settles the bet independently of any bettor's grasp of it. The proposition is an objective element of the agreement between bettors.

Resolving the problem of direct reference requires generalizing the principle of expected utility to accommodate probability's, and hence expected utility's, relativization to ways of grasping propositions. The reformulation has two steps. First, one assigns to each atomic state of an algebra furnishing partitions a way of grasping that state and assigns to compound states analogous compound ways of grasping them. Second, one demands invariance of an option's expected utility only with respect to partitions formed using the same algebra.

This reformulation of the expected utility principle recognizes the influence of framing effects on rational choice in cases such as the example. An option's expected utility is relative to a partition of states, whose probabilities are relative to ways of grasping them. Hence, a ranking of options according to expected utilities, and an option's having maximum expected utility, are relative to ways of grasping propositions. Ways of grasping propositions affect rational behavior, and expected utilities must take account of them to yield principles of rational behavior.

The relativization is conservative. If an agent makes a set of decisions using the same algebra of states relative to constant ways of grasping them, then standard representation theorems ensure that the agent's assignment of probabilities to the states may be inferred from the decisions, given satisfaction of the theorems' assumptions about the rationality of the agent and the structure of the set of decisions.

It is possible to individuate decisions according to frames. For example, one may distinguish a decision to bet that Cicero is Cicero from a decision to bet that Tully is Cicero because the decisions involve different ways of grasping their common content. In analogy with belief, decision may be a relation between an agent, time, proposition, and way of grasping the proposition. However, even making this reasonable move, a decision still needs relativization to a frame. For a decision so specified may nonetheless maximize expected utility with states framed one way and not maximize expected utility with states framed another way. For example, the decision to bet that Cicero is Cicero may maximize expected utility with respect to states presented as {Cicero is Cicero, Cicero is not Cicero} but not maximize expected utility with respect to the same states presented as {Tully is Cicero, Tully is not Cicero}.

Relativizing decisions to ways of grasping their contents and relativizing expected utilities to ways of grasping atomic states sanctions posting different odds for bets on the same proposition. Such postings invite a Dutch book, a combination of bets such that one loses money no matter what. However, falling victim to a Dutch book is not irrational, even for an ideal agent, if the necessity of losing money is not a priori but rather a posteriori. For example, suppose that an agent sells for $0.50 a gamble that pays $1.00 if Tully is Cicero and nothing otherwise. That wager may seem incoherent, but is rational if the agent's grasp of the proposition at issue prompts him to assign it a probability of 0.5. Ignorance of the empirical fact that Tully is Cicero excuses the agent's wagering in a way that ensures a loss of $0.50.

Given commonplace idealizations, a decision is rational if it maximizes utility with respect to the agent's representation of the decision problem, provided that his representation is acceptable. Two points about this standard of rationality are especially important. First, the standard is not utility maximization with respect to some frame or other. Rather, it is utility maximization with respect to the frame the agent actually uses, even if alternative frames are acceptable. A decision's rationality depends on the agent's way of framing his decision problem, that is, his way of grasping relevant propositions.

Second, a decision's evaluation relative to a frame, if comprehensive, considers whether the frame and its use are reasonable. Should the agent have used a different frame? Should the agent have noticed his frame's equivalence to other frames pointing toward different decisions? Should the agent have sorted out conflicts among recommendations arising from various frames before deciding? Rationality's tolerance of different decisions with respect to different frames presumes that an agent has made a reasonable effort to resolve inconsistencies among frames. In particular, it assumes that he has investigated the identity of propositions as much as his decision problem and his abilities warrant. Granting a reasonable effort to achieve coherence, remaining failings acceptable in the sense of Weirich (2004b, Chap. 6) do not discredit the decisions they influence.

Relativizing a decision's rationality to an agent's frame makes identifying that frame important. How does an agent frame her decision? Sometimes an agent's deliberations adopt a single frame. Then her frame is clear. If an agent frames a decision two ways, what matters is the frame her deliberations use to reach her decision. Sometimes an agent decides without deliberation. Then her frame is less clear. Perhaps the main evidence for a frame is the decision itself. The agent grasps the proposition that forms the content of her decision. The way she grasps it indicates the frame for her decision. If she takes herself to have bet that Cicero is self-identical, then the frame for her decision is clear. In general, a decision's rationality is relative to the way the agent frames her decision in deliberations yielding it if she consciously uses a particular frame, and is relative to the frame that is the best explanation of her decision if she does not consciously use a particular frame.

In cases where an agent's frame is indeterminate, her decision's rationality may also be indeterminate. For example, suppose an agent may frame her decision problem multiple ways, and the frame that steers her decision, if any, is indeterminate. If different frames support different decisions, an evaluation of her decision's rationality may lack an adequate metaphysical foundation. Evaluations should not move beyond the grounds for them.

Conclusion

Human rationality is "bounded." More precisely, the requirements of rationality are sensitive to human limits. Direct reference draws attention to a person's inability to identify propositions presented in different guises. Rationality recognizes that limit. This paper presents probability and decision principles that accommodate it. They retain the standard idealization of logical omniscience but dispense with knowledge of necessary truths, in particular, a posteriori necessary truths concerning identities of propositions. They apply to agents ignorant of such empirical matters.

Relativizing probabilities to ways of grasping propositions improves accounts of rational behavior because relativizing brings greater realism. This realism benefits economics and other behavioral sciences. It benefits philosophy, too. It makes normative decision rules psychologically realistic without sacrificing precision. It generalizes probability and decision principles so that they cover cases of incomplete information about identities of propositions.

My main conclusion is that a decision's rationality depends on the way the agent frames his decision problem. The decision's rationality is relative to that frame even if the agent is fully rational, cognitively ideal, and assigns precise probabilities and utilities to all propositions relevant to his decision problem. His frame is as significant as his information and goals in an evaluation of his decision's rationality. A comprehensive evaluation of his decision considers his way of understanding pertinent propositions.

Acknowledgements My colleagues Claire Horisk, Brian Kierland, Andrew Melnyk, and Peter Vallentyne made valuable suggestions, as did participants at the May, 2004 meeting of the Society for Exact Philosophy. I thank them all.

References

Aumann R (1976) Agreeing to disagree. Ann Stat 4:1236–1239

Binmore K (1994) Game theory and the social contract, vol 1. Playing fair. MIT Press, Cambridge, MA

Burge T (1979) Individualism and the mental. In: French P, Uehling T, Wettstein H (eds) Studies in metaphysics, midwest studies in philosophy, vol 4. University of Minnesota Press, Minneapolis, pp 73–121

Chalmers D (2002) The components of content. In: Chalmers D (ed) Philosophy of mind: classical and contemporary readings. Oxford University Press, New York, pp. 608–633
Crimmins M (1992) Talk about beliefs. MIT Press, Cambridge, MA
Debreu G (1959) Theory of value: an axiomatic analysis of economic equilibrium. Wiley, New York
Fodor J (1998) Concepts: where cognitive science went wrong. Clarendon, Oxford
Gaifman H (1988) A theory of higher order probabilities. In: Skyrms B, Harper W (eds) Causation, chance, and credence, vol 1. Kluwer, Dordrecht, pp 191–220
Garber D (1983) Old evidence and logical omniscience in Bayesian confirmation theory. In: Earman J (ed) Testing scientific theories. University of Minnesota Press, Minneapolis, pp 99–131
Gärdenfors P (1975) Qualitative probability as an intensional logic. J Philos Logic 4:171–185
Harsanyi J (1967) Games with incomplete information played by 'Bayesian' players. Part I: The basic model. Manage Sci 14:159–182
Horgan T (2000) The two-envelope paradox, nonstandard expected utility, and the intensionality of probability. Noûs 34:578–602
Jeffrey R (1983) The logic of decision, 2nd edn. Chicago University Press, Chicago, IL
Jeffrey R (1992) Probability and the art of judgment. Cambridge University Press, Cambridge
Joyce J (1999) The foundations of causal decision theory. Cambridge University Press, Cambridge
Joyce J (2002) Levi on causal decision theory and the possibility of predicting one's own actions. Philos Stud 110:69–102
Kahneman D, Tversky A (1979) Prospect theory. Econometrica 47:263–291
Kahneman D, Tversky A (1982) The psychology of preferences. Sci Am 246:160–173
Kaplan D (1989) Demonstratives. In: Almog J, Perry J, Wettstein H (eds) Themes from David Kaplan. Oxford University Press, Oxford, pp 481–614
Levi I (2000) Review essay: The foundations of causal decision theory. J Philos 97:387–402
Lewis D (1983) Attitudes *De Dicto* and *De Se*. In: Lewis D (ed) Philosophical papers, vol. I. Oxford University Press, New York, pp 133–159
McKinsey M (1999) The semantic of belief ascriptions. Noûs 33:519–557
Morris S (1995) The common prior assumption in economic theory. Econ Philos 11:227–253
Peterson M (2003) Transformative Decision Rules: Foundations and Applications. Theses in Philosophy from the Royal Institute of Technology 3. The Royal Institute of Technology, Stockholm
Resnik M (1987) Choices. University of Minnesota Press, Minneapolis
Rubinstein A (1998) Modeling bounded rationality. MIT Press, Cambridge, MA
Sahlin N (1994) On higher order beliefs. In: Dubucs J (ed) Philosophy of probability. Kluwer, Boston, MA, pp 13–34
Salmon N, Soames S (eds) (1988) Propositions and attitudes. Oxford University Press, New York
Savage L (1972) The foundations of statistics, 2nd edn. Dover, New York
Schick F (2003) Ambiguity and logic. Cambridge University Press, Cambridge
Skryms B (1980) Higher order degrees of belief. In: Mellor, DH (ed) Prospects for pragmatism: essays in memory of F. P. Ramsey. Cambridge University Press, Cambridge, pp 109–137
Skryms B (1984) Pragmatics and empiricism. Yale Unversity Press, New Haven, CT
Skryms B (1990) The dynamics of rational deliberation. Harvard University Press, Cambridge, MA
Stalnaker R (1984) Inquiry. MIT Press, Cambridge, MA
van Fraassen B (2002) The empirical stance. Yale University Press, New Haven, CT
Weirich P (1983) Conditional probabilities and probabilities given knowledge of a condition. Philos Sci 50:82–95
Weirich P (2001) Decision space: multidimensional utility analysis. Cambridge University Press, Cambridge
Weirich P (2004a) Belief and acceptance. In: Niiniluoto I, Sintonen M, Wolenski J (eds) Handbook of epistemology. Kluwer, Dordrecht, pp 499–520
Weirich P (2004b) Realistic decision theory: rules for nonideal agents in nonideal circumstances. Oxford University Press, New York

Epilogue

Propensities and Frequencies:
Inference to the Best Explanation*

James H. Fetzer

Perhaps no principle of reasoning holds more promise for understanding the foundations of scientific inquiry than that of *inference to the best explanation*. In its general form, this is a species of inductive inference that involves selecting one member from a set of alternative hypotheses as the hypothesis providing the best explanation for the available evidence. Alternatives that explain more of the available evidence are preferable to those that explain less, while those that are preferable when sufficient evidence is available are also acceptable. Acceptable hypotheses may still be false, making reasoning of this kind fallible, but remain the most rational of those alternatives under consideration.

This approach toward understanding science has often been associated with the name of Charles Peirce, whose conception of *abductive inference* properly qualifies as a mode of creative conjecture that might best be envisioned as a psychological process rather than as a logical procedure, which functions as a heuristic as opposed to an algorithm. Among the points that I shall address is that abductive inference complements inference to the best explanation as an essential component of a conception of scientific inquiry called "abductivism", where abductivism seems to supply a more adequate reconstruction of science than do its "inductivist", "deductivist" and "hypothetico-deductive" alternatives.

It has often been said, in the spirit of Occam, that simpler theories ought to be preferred to complex alternatives. That maxim, however, properly implies that simpler theories are preferable theories only when they are also adequate. If we assume that science aims at the discovery of *laws of nature* that have the form of general principles that are applicable for purposes of explanation and of prediction, then abductivism not only supplies a more complex theory of science than its inductivist and deductivist alternatives, but also appears to be the only account that is adequate.

J.H. Fetzer (✉)
McKnight Professor Emeritus, University of Minnesota Duluth, 800 Violet Lane, Oregon, WI 53575
e-mail: jfetzer@d.umn.edu

*An earlier version of this paper was presented to the Annual Meeting of The British Society for the Philosophy of Science held at The London School of Economics, London, UK, on 8 June 1998, and to the Department of Philosophy of The University of Glasgow, Glasgow, Scotland, on 9 June 1998. For another study of these important questions, see especially Fetzer (2000b).

E. Eells and J.H. Fetzer (eds.), *The Place of Probability in Science*, Boston Studies in the Philosophy of Science 284, DOI 10.1007/978-90-481-3615-5_14,

Establishing that this is indeed the case, however, not only presupposes establishing the conditions that scientific explanations have to satisfy but also entails establishing the limitations of these alternative accounts.

Measuring degrees of goodness of explanations, where some are better than others, requires distinctions between possible and actual explanations and between probabilistic and deterministic explanations, where differences between *probabilities as propensities* and *probabilities as frequencies* matter most. The analysis of science advocated here integrates a Popperian conception of natural laws with a modified Hempelian account of explanation, where Hacking's law of likelihood (in nomic guise) serves a crucial inferential function. These elements yield a coherent point of view that succeeds where its alternatives cannot, while clarifying and illuminating fundamental aspects of ontology and of epistemology.

The Inductivist Model

The place to begin, no doubt, is with the inductivist conception, which is both the simplest and the least adequate account of scientific inquiry. Drawing inspiration from Hume's reduction of causation to relations of resemblance, temporal succession and spatial contiguity, this approach envisions deterministic laws as no more than constant conjunctions between reference properties and attributes. Probabilistic (or, better, statistical) laws are identified with relative frequencies in finite sequences or as the limits of relative frequencies in infinite sequences. The most influential 20th C. representatives of this conception among philosophers, no doubt, have been Hans Reichenbach and his student Wesley C. Salmon.

This approach envisions scientific inquiry as a process or procedure involving four stages or steps, which are supposed to be followed in this sequence, namely (see Fig. 1):

Observation

Classification

Generalization

Prediction

Fig. 1 The four stages of inductivism

where the fundamental principle of inference is "the straight rule", namely: if m/n observed As are Bs, then infer (inductively) that m/n As are Bs, provided that a suitable number of As have been observed over a wide variety of conditions (see Reichenbach 1937, 1949; Salmon 1967, 1971).

A schematization of this approach is easily supplied, where the double-line between premises and conclusion indicates that such an inference is inductive:

1a] m/n As have been observed to be Bs.
The sample of n is large and varied.
======================
m/n As are Bs.

The definition of probabilities as relative frequencies, moreover, justifies the translation of conclusions of the form, 'm/n As are Bs', into logically equivalent probabilistic hypotheses of the form, '$P(B/A) = m/n$', that might function as premises in inductive explanations having more or less the following form:

1b] $P(B/A) = m/n$.
x is an A.
=========== [m/n]
x is a B.

where the bracketed value '[m/n]' is a logical probability that represents the truth frequency with which conclusions of that form are true when premises of that form are true, as Reichenbach (1949) proposed. This approach might incorporate a condition that [m/n] must equal or exceed some specific value, such as 1/2, for example, for explanations or for predictions to be adequate.

Thus, such an approach could justify inferring from the premise that three fourths of the rabbits in a sample have been observed to be white to the conclusion that three fourths of all rabbits are white, from the premise that drinks of water have been observed to quench thirst to the conclusion that water quenches thirst, from the premise that bodies in free fall have been observed to accelerate at 32 fps/ps to the conclusion that bodies in free fall accelerate at 32 fps/ps, and from the premise that locations of planets orbiting around the Sun have been observed to be points on elliptical paths to the conclusion that the orbits of the planets are elliptical, together with corresponding explanations and predictions.

Inductivist Inference

Without doubt, inductivism provides a simple model of scientific inquiry that combines a Humean conception of natural laws as constant conjunctions or as relative frequencies with a principle of inference that appears to accommodate them. Moreover, when more evidence becomes available, conclusions that were previously justified may be replaced by alternative conclusions in conformity with *the requirement of total evidence*. When science is assumed to aim at the discovery of relative frequencies or their limits within sequences, the straight rule can be vindicated pragmatically by the argument that it has the capacity to discover those values, if any rule can, provided only they exist.

The inductivist conception, however, encounters certain objections. When interpreted algorithmically, it specifies that scientific inquiries have to begin with observation. If this injunction means that scientists must observe everything that could possibly be observed before proceeding to classification, then it would never reach that stage, since there is an infinity of things that might be observed. Every single event, such as the first time that Galileo dropped bodies in free fall from the Tower of Pisa, has innumerable relations to other events, such as its distance from the Moon, the Sun and other celestial objects, which could never be exhaustively described. Thus, as Carl G. Hempel (1966) remarked, scientific inquiries, thus conceived, could never get off the ground.

The reply that Reichenbach and Salmon would advance, no doubt, is that this criticism involves attacking a straw man, because scientists never have to observe everything that could possibly be observed before proceeding to classification. What needs to be observed, they would say, does not include everything but only what is relevant. The relevance of observations (experiments, measurements), however, depends upon the prior specification of one or more hypotheses with respect to which those observations become relevant. If observations only become relevant in relation to specific hypotheses, then inductivism is not right. The stages of inductivism occur in the wrong order.

Moreover, although inferences from samples to populations are important, inferences of this kind appear to be confined to the observable properties of observable entities, where the only sense in which they allow inferences from the observed to the unobserved is when the "unobserved" has simply not yet been observed! Inferences from observations about falling bodies might permit inferences about other falling bodies, but would not permit inferences to non-observable properties, such as gravitational attraction and relative mass. The achievements of Newton transcended those of Galileo and of Kepler, not merely by their subsumption, but by subsuming them by means of a theory.

Inductivist Explanation

Somewhat surprisingly, the inductivist account implies that the conditions for predictions do not necessarily coincide with those for explanations. Since probabilities for attributes vary in different reference classes, singular events need to be assigned to appropriate reference classes. When other properties make a difference to the outcome – such as when water is salty as opposed to fresh, for example – their existence can be accommodated through the supplemental principle of basing predictions upon *the narrowest reference class for which reliable statistics are available*, as Reichenbach recommended, a standard that applies in relation to the current state of our statistical knowledge.

The conception that a property is statistically relevant to the occurrence of another property when its presence or absence makes a difference to the relative frequency with which that property occurs, however, could apply as an ontic rather than as an epistemic measure by imposing the more stringent condition of truth. Salmon therefore proposed that the occurrence of singular events should be explained by assigning them to *the broadest homogeneous reference classes to which they belong* by taking into account every property of these events that is statistically relevant to their attributes (Salmon 1971).

The statistical relevance criterion combined with the requirement of ontic homogeneity, however, has the uningratiating consequence of making statistical explanations of singular events logically impossible, in principle. Attributes can occur with relative frequencies relative to *some* of the properties of their predecessors when those predecessors are unique, but since explanations of their occurrence are required to take into account *every* property whose presence affects their relative frequency, there is no theoretical latitude for discounting conjunctions of properties instantiated by these events, which may or may not have other instances, relative to which those attributes occurred (Fetzer 1981).

Thus, in any case in which the relative frequency for an attribute B is not equal to one or to zero, there will always exist additional properties F of their predecessors A that qualify as "statistically relevant", on pain of contradiction. The conditions for adequate predictions, it turns out, are not necessarily the same as those for adequate explanations, because the result that every event occurs with a degenerate probability of one or zero does not preclude predictive adequacy. Every event occurs or fails to occur, which is exactly what we want to know. But it vitiates the statistical relevance theory of explanation.

Inductivist Regularities

While the implicit restriction to observable properties delimits the kinds of hypotheses within the scope of inductivist procedure, explicit reliance upon the straight rule itself raises other concerns. The large number and wide variety conditions are intended to insure that samples are "random", at least in the epistemic sense that measures have been taken to guard against bias. The application of this rule is equivalent to the decision to treat such samples as *representative*, where the relative frequencies of properties within those samples are supposed to correspond to those of the population entire. Yet this approach supplies no ontic rationale to explain why that should be true.

The absence of an ontic rationale for the presumptive representativeness of epistemically random samples, which are *believed* to be unbiased because they reflect a suitable number observed under a wide variety of conditions, is part and parcel of the Humean character of the conception of laws as merely constant conjunctions or relative frequencies. Another ontic objection to the inductivist account thus becomes the failure to differentiate between causation and correlation. Any two properties, however arbitrarily selected, such as shapes and colors, like round and red, occur with some relative frequency during the history of the world. But they are not therefore lawfully related.

Additional conditions have therefore been introduced, including the identification of lawful regularities as those for which m/n converges to a limit p. Conditions of insensitivity to place selection and of freedom from aftereffect have also been proposed to distinguish laws from accidental generalizations, where a sequence of instances of As may be said to be *normal* when its outcomes B (non-B, and so on) have relative frequencies that neither depend upon those of their predecessors nor differ from those for subsequences selected by taking every kth instance (Reichenbach 1949, pp. 143–144). These sequences of As have to be infinite, however, since otherwise they cannot be satisfied for large k. Without infinite sequences, therefore, laws of nature cannot exist.

The existence of infinite sequences within abstract contexts might pose no special problems, but the existence of infinite sequences in physical contexts is another thing. If there were an ontic rationale for epistemic randomness, then it might be appropriate to view lawful regularities as *subjunctive conditionals*, for which probabilities specify the limits with which specific attributes would occur if their reference sequences were infinite, and accidental generalizations turn out to be regularities without subjunctive counterparts. Unfortunately, the inductivist approach affords no foundation for drawing such a distinction, apart from Hume's "habits of mind", which are psychological rather than ontological.

The Deductivist Model

By insisting that laws have *the force of prohibitions*, therefore, Popper (1965) captured a crucial difference between natural laws and merely accidental regularities. Unlike laws of society, for example, which can be violated and can be changed, laws of nature cannot be violated and cannot be changed. Rather than begin with observation, moreover, deductivism envisions the introduction of hypotheses and theories as a psychological process rather than as logical procedure, where their credibility depends upon their ability to withstand our best efforts to refute them: they are created or contrived as guesses or conjectures that are not *derived from* experience, but instead *tested by* experience.

Popper's emphasis upon corroboration as distinct from confirmation, moreover, reflects the difference between conceptions of laws as regularities with an ontic foundation and as regularities without an ontic foundation. While confirmation searches for positive instances of hypotheses, the only evidence that should count in favor of the existence of *lawful* regularities arises when we seriously attempt to falsify, refute or disconfirm them and do so without success! Unsuccessful attempts to violate presumptive regularities must therefore be distinguished from successful attempts to find positive instances. Science thus proceeds through the process of forming conjectures and attempting to refute them.

When understood as implementing corroboration in lieu of confirmation, this approach likewise envisions scientific inquiry as a process or procedure involving four steps or stages, which are supposed to be followed in this sequence, namely (see Fig. 2):

Conjecture

Derivation

Experimentation

Elimination

Fig. 2 The four stages of deductivism

where the fundamental principle of inference is "modus tollens" in application to hypotheses as follows: if hypothesis H is true, then its logical consequence C must also be true; but (as observations, measurements or experiments show) it is not the case that C is true; thus, infer (deductively) that hypothesis H is false.

The advantages of deductivism over inductivism are conspicuous relative to the simplest case. Even if 3/4 of rabbits have been observed to be white, that relative frequency need not be lawful. The percentage of rabbits that are white could be subject to manipulation, say, by placing a bounty on non-white rabbits and hunting them down. There are underlying laws that relate rabbits to their color, but they concern unobservable properties of rabbit genes. If white W is a dominant gene and non-white n is a recessive, then if these genes were equally distributed and randomly combined, about 3/4 of all rabbits *would be* born white. Conjectures about phenomena thus not only permit the introduction of theoretical hypotheses about non-observable properties but also guide inquiry.

Deductivist Inference

The fundamental rule of deductivist reasoning may likewise be schematized in application to hypotheses H and their consequences C, where the single line between premises and conclusion indicates that such an inference is deductive:

6a] If H is true, then so is C.
But (as the evidence shows) C is not true.
———————————————
H is not true.

Indeed, this pattern of reasoning commonly occurs within abstract contexts, where arguments depend upon formal methods rather than on observation, measurement, or experiment. Within physical contexts of the kind that distinguish empirical science, its application tends to require a somewhat more complex formulation, as many students, including Popper, have pointed out.

Since the derivation of testable consequences from theoretical hypotheses typically depends upon additional premises, this naive conception must be displaced by a more sophisticated version, as even the most classic cases display. The Bishops of Padua, for example, sought to deflect the falsifying significance of Galileo's telescopic observations of the irregular and pockmarked surface of the Moon – which if accepted refuted the Aristotelian doctrine that all celestial objects are perfect spheres – by objecting that Galileo must have been pointing his telescope at some object other than the Moon or else that spherical objects simply do not look spherical when they are observed by means of a telescope.

Indeed, the Bishops were on firm ground, insofar as the truth of auxiliary hypotheses A concerning experimental apparatus (such as telescopes, microscopes, and the like) and initial conditions I concerning test subjects (such as that the object under observation is the Moon) are invariably assumed in the process of deriving testable consequences C from any testable hypothesis H. The naive deductivist model must therefore be displaced by its sophisticated counterpart, which takes into account the role of additional factors as follows,

6b] If H and A and I are all true, then so is C.
But (as the evidence shows) C is not true.

H and A and I are not all true.

where the weight of the evidence against H being true depends on the strength of collateral evidence for the truth of A and I (cf. Hempel 1966, p. 7 and p. 23).

Ironically, these considerations even extend to the outcome of "naked-eye" observations, insofar as things that are green, for example, typically look green under standard conditions of lighting to individuals who possess normal vision, but not in the dark when lights are off or when suffering from color-blindness. While the advantages of deductivism over inductivism include the freedom to create hypotheses and theories, unconstrained by their direct or indirect connections to observation and experience, as classical Newtonian mechanics illustrates, even naked-eye observations turn out to be "theory-laden" in ways that undermine the plausibility of inductivism while enhancing that of deductivism.

Deductivist Regularities

Popper sought to extend the scope of deductivist methodology to encompass empirical tests of probabilistic hypotheses by embracing the principle that improbable results, although logically possible, should be presumed to not occur. Even as an advocate of the conception of probabilities as limiting frequencies, Popper extended the principle of falsifiability to statistical conjectures on the basis of the supplemental principle that, insofar as almost all finite segments of large size will exhibit relative frequencies that are very close to the limiting frequencies of these infinite sequences, highly improbable outcomes deserve systematic neglect (Popper 1965, pp. 410–419). The methodology of conjectures and refutations was thereby extended far beyond universal hypotheses.

While he thus sought to take advantage of classic limit theorems that apply to sequences that are "normal" in the frequentist sense, Popper's subsequent introduction of the propensity interpretation of probability affords a much firmer foundation for his own methodological maxims. While propensities may be envisioned as dispositional properties of arrangements of conditions (as "chance set-ups") to bring about specific outcomes with limiting frequencies that equal their long-run propensities over infinite sequences of trials, as Popper initially envisioned them, they may also be construed as tendencies to bring about the same outcomes with specific strengths on singular trials (Popper 1959, 1990; Fetzer 1981, 1993).

The adoption of the single-case conception brings with it many advantages, including, for example, an ontic account of random sequences. Thus, when a sequence consists of replications of the same relevant conditions, where each trial has propensities of the same strength for each of the same possible outcomes and the outcomes of no trials affect the strengths of the outcomes for any other, they assume the character of independent, identically distributed random variables. Even though these are physical and not merely abstract sequences, classic mathematical results, such as the central limit theorem, become applicable for inferring from propensities to frequencies (Gnedenko and Kolmogorov 1954 and Fetzer 1981).

The propensity account of randomness thus probabilistically implies that, when H is true, almost all sequences of large size will exhibit relative frequencies that are very close to their generating propensities. It thereby provides the ontic rationale for epistemic randomness that frequentist interpretations are unable to supply. It relinquishes a definition in terms of infinite sequences for a definition in terms of singular events – "the single case" – where relative frequencies in finite sequences – "the short run" – as well as limiting frequencies in infinite sequences – "the long run" – are appropriately viewed as finite and infinite sequences of singular events during the world's history.

Permanent Properties

Thus, unlike probabilities as frequencies, probabilities as propensities can exist no matter whether the history of the world is short or is long. When the conception of laws as prohibitions receives appropriate elaboration, moreover, then the differences between different kinds of laws becomes apparent. Thus, an attribute A is a *permanent property* of everything that has a reference property R when, even though the possession of A is not part of the definition of R, there is no process or procedure, natural or contrived, by means of which something that possesses R could lose A, except by losing the reference property R. Attributes things could lose and remain R are merely transient (Fetzer 1977, 1981).

When "gold" is defined by means of its atomic number, which is 79, things that are gold have many permanent attributes, including being a yellow, malleable metal, which has a melting point of 1064°C and a boiling point of 3080 °C (Ruben 1985, p. 35). These are properties that things that are gold could lose only by losing their atomic number, something that could be done by nuclear bombardment, for example; otherwise, these relations between properties are invariable: they cannot be violated and cannot be changed. Other properties of things that are gold, however, such as their size, their shape, or their selling price, could be taken away from them without affecting their atomic number.

The introduction of non-logical necessary connections of this kind entails the conception of *lawlike sentences* as subjunctive conditionals whose consequents are dispositional predicates (Fetzer 1981). Indeed, the perspective afforded by the single-case propensity conception suggests that dispositions should be viewed as *single-case causal tendencies* of variable strength, where those that invariably bring about the same outcome under the same conditions (such as a melting point of 1064 °C) are deterministic, while those that bring about one or another outcome within a fixed class of possible outcomes with constant probabilities (such as a half-life of 3.05 min) are indeterministic.

Popper endorsed the general conception of explanation by subsumption but did not otherwise pursue it. While Hempel's more detailed covering-law model would confront criticisms of the kind that Salmon and others advanced, a Popperian approach toward understanding the nature of laws of nature – inspired by the conception of laws as prohibitions and the propensity account of probability – provides crucial ingredients for overcoming them. Most importantly, the difference between universal and probabilistic laws is not a matter of how many members of a reference class possess the attribute of interest but of the comparative strength of the attribute possessed by every member of that class.

Hypothetico-Deductivism I

The conditions for Popperian rejection were always more conspicuous than the conditions for Popperian acceptance. Popper distinguished between acting on an hypothesis as though it were true and accepting that hypothesis as true. While we are entitled to act on the basis of well-corroborated theories, namely: those that have survived our best attempts to refute them, we must recognize that even our best theories may turn out to be false. They are forever vulnerable to new observations, measurements, and experiments or to the invention of new theoretical alternatives. Our failure to discover falsifying conditions is no guarantee they do not exist; we may have not looked in the right places.

Thus, although Newton's and Einstein's theories of gravitation are strictly inconsistent, they yield the same results for weak gravitational fields and for objects moving slowly (in relation to the speed of light). Tests within their overlapping domain of application could not possibly discriminate between them; but beyond this range of overlap, they offer incompatible outcomes, which provide opportunities to differentiate between them by severe observational and experimental tests, which have corroborated Einstein while refuting Newton. Since Newton's theory was surely the most thoroughly confirmed in the history of science, this is a stunning example of Popper's point.

Of the three most common forms of inductive reasoning – from samples to populations, from the observed to the unobserved, and from the past to the future – therefore, the deductivist conception most conspicuously exemplifies the pattern of drawing inferences from the observed to the unobserved. But it does so by inviting conjectures about the unobservable, the non-observable, and the theoretical, which are subjected to empirical tests on the basis of their deductive consequences. It thereby promotes the invention of hypotheses and theories of broad scope and systematic power. What it fails to do, however, is to establish suitable conditions for the acceptance of hypotheses and theories.

The difference between hypothetico-deductivism and deductivism within this context, therefore, is that hypothetico-deductivism likewise envisions scientific inquiry as involving four steps or stages, three of which are the same (see Fig. 3):

Conjecture

Derivation

Experimentation

Acceptance

Fig. 3 The four stages of hypothetico-deductivism

where the fundamental principle of inference in application to hypotheses assumes something more or less like the following form: if hypothesis H is true, then its logical consequence C must be true; as observations, measurements or experiments show, C is true; thus, infer (inductively) that hypothesis H is true.

Hypothetico-Deductivism II

While hypothetico-deductivism thus promises to go beyond deductivism by supporting the acceptance of hypotheses and theories, whether it can deliver is another matter. Were this principle of inference construed deductively, it would display the pattern characteristic of the following argument form,

10a] If H is true, then so is C.
(As the evidence shows) C is true.
===================
H is true.

which is a familiar fallacy known as "affirming the consequent", where the truth of the premises provides no guarantee of the truth of the conclusion. Consider: if Sarah is a freshman, then she is a student, and she is a student; or if Steven won the lottery, then he bought a ticket, and he bought a ticket.

Indeed, the suggestion that deductively invalid arguments should provide the foundation for inductive reasoning led to the witticism that books on logic were divided into two parts, the first on deduction (where the fallacies were explained), the second on induction (where they were committed). Yet it is almost impossible to deny that, if more of its deductive consequences can be shown to be true, the greater the reason for accepting the hypothesis as true:

10b] If H is true, then so is C1, C2, . . . , Cn.
(As the evidence shows) C1, C2, . . . , Cn are all true.
================================
H is true.

If Sarah is a freshman, then she is a student, lives in the dorms, takes introductory classes, and has not declared a major; Sarah is a student who lives in a dorm, takes introductory classes and has not declared a major. It does not guarantee that Sarah is a freshman, but the argument has grown far stronger.

It would be wrong to assume that arguments that are deductively unsound are therefore inductively incorrect. Every inductive argument that has a proper form must be deductively invalid, precisely because its conclusion contains more content than do its premises. The failure to satisfy deductive standards is not at fault. Nevertheless, it does seem rather odd that a form of argument that is deductively defective should turn out to be inductively virtuous if only it is committed often enough. If these patterns of argument are the best to be found on behalf of acceptance, perhaps Popper had the right attitude, after all.

The problem of acceptance, however, has to be resolved; otherwise, science cannot attain the aim of discovering general principles that can be applied for the purposes of explanation and of prediction. Indeed, even if *predictions* can be based upon mere correlations in conformity with inductivist procedures, in order to be adequate, *explanations* must be true. Whether or not the covering-law account might require modification, as long as explanation occurs by subsumption, explanations cannot be adequate in the absence of premises that can function as covering laws. The success of science thus depends upon its ability to fallibilistically accept hypotheses and theories as true (Fetzer 2000a, p. xxvi).

The Abductivist Model

Like its inductivist and deductivist alternatives, the abductivist model can be described as characterizing scientific inquiries as having four steps or stages (see Fig. 4):

Puzzlement

Speculation

Adaptation

Explanation

Fig. 4 The four stages of abductivism

The fundamental principle is "inference to the best explanation", which involves selecting one member from a set of alternative hypotheses as the hypothesis that provides the best explanation for the available evidence. Alternatives that explain more of the available evidence are *preferable* to those that explain less, while those that are preferable when sufficient evidence is available are also *acceptable*. Acceptable hypotheses may still be false, which makes reasoning of this kind fallible, but they are the most rational among the alternatives.

Unlike inductivism and deductivism, abductivism does not presume observations or conjectures must come first but rather assumes instead that human curiosity as a psychological phenomenon motivates scientific inquiries. Here, no doubt, a distinction can be drawn between *pure* and *applied* science, since some discoveries of hypotheses and theories have more mundane motivation. The process of adaptation, moreover, which shall be considered in some detail, incorporates deductivist principles on the basis of incorporating the procedure that hypotheses that are incompatible with the available evidence are rejected as false, where rejection reflects a tentative and fallible status that can change.

An important desideratum for abductivism is that every relevant alternative receive consideration, since otherwise the true hypothesis may not be included among those under investigation. Peirce proposed a form of thinking he called "abductive inference" that might serve an heuristic function, namely:

11a] The surprising fact C is observed.
If H were true, then C would be a matter of course.
- -
There is reason to suspect that H is true.

where the broken line between premises and conclusion is meant to indicate that Peirce did not consider this to be a form of inference that would permit drawing H as a conclusion, but only as a speculative mode of creative conjecture that H might deserve consideration (Hartshorne and Weiss 1935, 5.189).

Abductive inference of this kind has a function at the stage of speculation but, unlike hypothetico-deductivism, does not determine which hypotheses or theories should be accepted. If abductivism commits the fallacy of affirming the consequent, therefore, this is not the stage of its occurrence. Presumably, Pierce would also embrace the hypothetico-deductive abductive counterpart:

11b] The surprising facts C1, C2, . . . , Cn are observed.
If H were true, then C1, C2, . . . , Cn would be a matter of course.
- -
There are multiple reasons to suspect that H is true.

Since there is no mechanical procedure to guarantee that such a process will generate every relevant alternative possible explanation H1, H2, . . . , Hn, abductivism always remains vulnerable to the discovery of new alternatives.

Abductivist Explanation I

Since every alternative must qualify as a possible explanation, the tenability of abductivism strongly depends upon the adequacy of the theory of explanation that provides its foundation. On initial consideration, the most promising account would appear to be the covering-law account Hempel has proposed, which implements the conception of explanation by subsumption. Hempel, of course, distinguished two basic models of explanation, one model for the deductive subsumption of phenomena under laws of universal form the other for their inductive subsumption under laws of probabilistic form.

Hempel envisioned explanations as arguments with premises that include sentences describing appropriate laws and initial conditions, together known as *the explanans*, and a conclusion describing the phenomenon to be explained, known as *the explanandum*. The four conditions that he advanced for the adequacy of deductive-nomological explanations are familiar to us all, namely:

(CA-1) the explanandum must follow from the explanans;
(CA-2) the derivation must appeal to some covering-law;
(CA-3) the explanans must have empirical content; and,
(CA-4) the sentences that constitute the explanans must be true.

Hempel observed that (CA-3) was implicitly satisfied by any explanans that satisfied the other conditions but was of sufficient interest to render explicit.

Although these conditions were intuitively appealing, other students located certain difficulties with Hempel's formulations. Among the most interesting were counterexamples advanced by Salmon (1971, p. 34), such as the following:

12a] Every man who takes birth control pills regularly avoids pregnancy.
John Jones regularly took birth control pills during the past year.
―――――――――――――――――――――――――――――――
John Jones avoided becoming pregnant during the past year.

Salmon would contend, with considerable justification, that Hempel's conditions were undermined by an inappropriate conception of explanatory relevance, but the statistical-relevance conception he proposed confronted problems of its own.

The birth control pill example, like others, satisfies Hempel's four conditions, yet fails to provide an adequate explanation. The argument itself exemplifies the logical form, *modus ponens*. Other counterexamples that have the logical form, *modus tollens*, are equally persuasive and may be even more revealing:

12b] Everyone who has ever been run over by a steam roller is dead.
Mary Smith is not dead.
―――――――――――――――――――――――――――――――
Mary Smith has never been run over by a steam roller.

Similar arguments could be fashioned in relation to other conditions that are sufficient to bring about death, such as being stepped on by an elephant, hit by a train, and so on (Fetzer 1992, p. 258). As long as someone remains alive, sufficient conditions for her (or his) death cannot, as yet, have been realized.

Abductivist Explanation II

It should be observed that the pattern of inference in the steam roller case works perfectly successfully for the purpose of *retrodiction*, just as the pattern in the birth control case works successfully for the purpose of *prediction*. Men do not become pregnant, whether or not they take birth control pills, which makes this *explanation* defective by introducing irrelevant conditions. And while the inference *that* anyone who is still alive has never been run over by a steam roller may be sound, it fails to explain *why* she (or he) has never been run over by a steam roller as an historical phenomenon.

These difficulties suggest that there must be more to explanation than derivation from premises that include universal laws. Other problems, however, were generated by Hempel's commitment to the underlying desideratum that explanations explain by displaying the *nomic expectability* of their explanandum events, which dictated that every adequate explanation must qualify as a potential prediction, had its premises been taken into account at a suitable time. This led him to impose the requirement that the logical probability between explanans and explanandum in the case of inductive-statistical (later, probabilistic) explanations must have a value equal to or greater than 1/2.

This condition had the disturbing effect that inductive explanations for phenomena that occur with low probability are logically impossible, in principle. If the outcome of an ace given a toss of a die occurs with a probability of only 1/6, for example, then such an outcome was not to be expected and therefore could not be explained. Unfortunately, it might be the case that each of the possible outcomes only occurs with a probability less than 1/2, as with tosses of dice, in general, where neither the outcome of an ace nor a deuce nor a trey could be explained, while the occurrence of a non-ace or a non-deuce or a non-trey could be explained, consistent with this account.

Moreover, Hempel made the adequacy of I-S explanations *relative to a knowledge context K*, thereby abandoning the truth condition imposed upon D-N explanations. His motives, I am convinced, were rooted in an apparent effort to forestall the consequence that the outcomes of unique events must have degenerate probabilities equal to zero or equal to one, which would follow from taking into account every property of the unique predecessor of any explanandum event. The result was virtually a schizophrenic theory of explanation, which mixed together conditions of adequacy of different kinds.

Abductivist Regularities I

In retrospect, the problems that afflicted Hempel's account seem to have had two important sources. The first is that his semiformal explication, (CA-1) to (CA-4), while intuitively enticing, lacked a requirement that would exclude the presence of irrelevant factors from the lawlike premises that may appear in an explanans.

The other was that the extensional methodology to which he was philosophically committed precluded adopting a conception of probability that could function adequately within the context of probabilistic laws. As a consequence, it was impossible to formally vindicate his theory of explanation.

These difficulties were most acute in the case of I-S explanations, where he attempted to fashion a condition that would establish an appropriate relationship between statistical premises and explanandum events. His initial formulation of *the requirement of maximal specificity* was envisioned as an epistemic condition implementing the total evidence requirement (Hempel 1962). Later Hempel recognized that establishing a suitable connection between events to be explained and probabilistic lawlike premises was a matter of subsumption rather than of inference, with a distinct rationale, even though "probabilistic explanations" would remain inconclusive kinds of arguments (Hempel 1968).

Thus, although Hempel flirted with the propensity interpretation in 1965, its adequate elaboration would entail the utilization of intensional methods at variance with his extensional commitments. Thus, with reservations, he adhered to a hypothetical long-run interpretation, where the probability of B, given A, equals its limiting frequency p within an infinite sequence, that is, $P(B/A) = p$. Hempel acknowledged that lawlike sentences, but not accidental generalizations, support subjunctive and counterfactual conditionals, and he described their testable implications for sequences of equal and independent trials, but he was never able to provide an ontic justification for this position.

Indeed, it should be apparent by now that the fundamental inadequacy of any frequency-based interpretation – whether short-run accounts based upon relative frequencies or long-run accounts based on limits – is its incapacity to establish an appropriate relationship between explanans and explanandum in probabilistic explanations. Accounts based upon relative frequencies attempt to subsume the occurrence of singular events by means of short-run generalizations, while accounts based upon limits attempt to subsume them by means of long-run generalizations. Neither, however, can successfully explain precisely how and why short-run or long-run generalizations apply to singular events.

Abductive Regularities II

The fundamental advantage of the deductivist conception of natural laws as subjunctive conditionals that attribute dispositional attributes B of constant strength to everything possessing the reference property A, therefore, is that these are single-case generalizations that subsume singular events when their initial conditions instantiate their antecedents. They are capable of supporting counterfactual conditionals because conditionals of that kind are subjunctives with false antecedents. Since subjunctives imply but are not implied by material conditionals, oddities such as the alleged paradoxes of confirmation do not arise and the role of idealization in science can be successfully explained.

Simple subjunctives of the form, '$p \Rightarrow q$', attributing permanent property relations, which are logically contingent, would be vulnerable to paradoxes of confirmation were they transposable. Like other intensional conditionals, including causal conditionals of universal and of probabilistic strength, however, they are not, as the steamroller case already exemplifies. The truth of '$p \Rightarrow q$' implies that both '$p \rightarrow q$' and '$-q \rightarrow -p$' is true, but does not imply $-p$ is a permanent property of $-q$, which would conflate logical and nomological necessities and thus contradict the assumption that $-\Box(p \rightarrow q)$ (Fetzer 1981, pp. 193–194).

That malleability M is a permanent property of things that are gold G justifies the singular subjunctive, '$Ga \Rightarrow Ma$', which in turn implies (the logically equivalent) subjunctive generalization, '$(x)(Gx \Rightarrow Mx)$', which is a lawlike sentence. '$(x)(Gx \Rightarrow Mx)$' implies its material counterpart, '$(x)(Gx \rightarrow Mx)$', which is logically equivalent to '$(x)(-Mx \rightarrow -Gx)$' and '$(x)(-Gx \vee Mx)$'. While observations of white shoes instantiate both the antecedent '$-Mx$' and the consequent '$-Gx$' of one of these material generalizations and, by Nicod's criteria, thereby confirm both it and its logical equivalents, none of them is a lawlike sentence.

Perhaps even more strikingly, the subjunctive conditionality of lawlike sentences means that natural laws make assertions about what would be the case, whether that happens to be the case or not. The role of *abstraction* in science thus seems to be that of specifying conditions that are and may always remain counterfactual. The influence of air resistance on the fall of light and heavy objects (feathers and cannon balls, for example), like the influence of friction on inclined plane experiments, is simply disregarded in formulating idealized laws for those circumstances. The claim that the laws of physics lie thus only applies to cases of subjunctive conditionality that exclude relevant conditions.

Abductivist Regularities III

The paradoxes of confirmation thus appear to arise from failing to distinguish direct tests of lawlike sentences, which require instantiations of their antecedents, from indirect tests, which do not. White shoes are red herrings relative to the hypothesis that gold is malleable as well as that ravens are black, which is reinforced by observing that there are no paradoxes of corroboration. The subjunctive conditionality of lawlike sentences combined with the desideratum that natural laws are described by lawlike sentences that are true thus entails that every predicate describing a property that makes a difference to the attribute's occurrence or its negation be implied by their antecedents.

The conception of relevance that applies here, however, is not statistical relevance, which Salmon mistakenly identified with explanatory relevance, but causal relevance or, more generally, nomic relevance. A property F may be said to be *causally relevant* to the occurrence of an attribute B, given the reference property A, when the strength of the tendency for B, given A & F, differs from the strength of the tendency for B, given A & $-$F. Then, employing the probabilistic causal conditional, ‘p =n⇒ q’, F is causally relevant to the occurrence of B, given A, when (x)[(Ax & Fx) =n⇒ Bx] and (x)[(Ax & − Fx) =m⇒ Bx], where the strength of these propensities vary and m $\neq$ n (Fetzer 1981).

Analogous definitions are available for nomic relevance relative to law-like sentences of subjunctive form, ‘(x)(Ax ⇒ Bx)’, and for causal conditionals of universal strength, ‘(x)[(Ax & Fx) =u⇒ Bx]’, employing the causal conditional of universal strength ‘p =u⇒ q’. The truth of lawlike hypotheses of any of these forms, however, clearly depends upon specifying some complete set of nomically or causally relevant factors in relation to the outcome phenomenon. Even the behavior of falling bodies cannot be successfully explained without taking into account appropriate frames of reference, the presence or absence of air resistance, and the relative masses of objects in gravitational interaction.

Thus, the appropriate condition to attain this relationship (which provides a basis for relating complete sets of initial conditions to explanandum outcomes), may also be called *the requirement of maximal specificity* (Fetzer 1981, p. 50). But it must be borne in mind that this requirement neither implements the requirement of total evidence nor attempts to locate singular events in relation to reference classes under a frequency interpretation of probability. Instead, it is an ontic condition for permanent property relations or single-case causal tendencies to obtain between reference properties and attributes in the world.

Abductivist Explanations III

The causal relevance (or nomic relevance) account of explanation that emerges from the propensity perspective thus distinguishes between simple laws of subjunctive form and causal laws of more complex forms, which may be deterministic or indeterministic in kind. There are therefore at least three kinds of explanations in science, namely: those whose premises include simple laws, those whose premises include deterministic laws, and those whose premises include indeterministic laws. These conceptions already overcome the problems generated by identifying statistical relevance with explanatory relevance. But they do not resolve the problems encountered with respect to irrelevant factors.

The solution to these difficulties is to impose an additional requirement, not on the truth of lawlike sentences, but upon the adequacy of explanations, namely: that the antecedents of the lawlike premises that occur in adequate explanations must exclude any properties whose presence or absence makes no difference to the occurrence of the explanandum event (Fetzer 1981, pp. 125–126). This condition, known as *the requirement of strict maximal specificity*, requires that adequate explanations may invoke as explanatory only properties whose presence made a difference to the explanandum event that is to be explained.

When these conditions are combined, the modified Hempelian account of explanation that emerges supports the following four conditions of adequacy:

(CA-1) the explanandum must be a deductive or probabilistic consequence of the explanans;
(CA-2) the explanans must include at least one lawlike sentence required for the derivation of the explanandum;
(CA-3) the explanans must satisfy the requirement of strict maximal specificity; and,
(CA-4) the sentences constituting the explanation – both the explanans and the explanandum – must be true.

where sets of sentences that satisfy the first three conditions are *possible* explanations and those that satisfy all four are *adequate* (Fetzer 1981, p. 127).

These conditions obviously resolve the problems posed by the birth control example, because the antecedent of the lawlike sentence that appears in the explanans does not exclude causally irrelevant conditions and fails to satisfy (CA-3). The steam roller example turns out to be even more instructive, because it becomes apparent that the entire explanans is causally irrelevant to the explanandum in violation of (CA-3). Indeed, being run over by a steam roller is insufficient to bring about death, since in marshy wetlands and other circumstances, it may cause injury but not death, as exemplified by a recent case near London (Jepson 1998), reflecting the importance of maximal specificity considerations

Nomic Expectability

Salmon abandoned the requirement that the logical probability between explanans and explanandum must equal or exceed 1/2, which was necessary to overcome the logical impossibility of explaining events that occur only with low probability. This was essential and marked the abandonment of Hempel's desideratum of nomic expectability for an alternative conception of nomic responsibility (Fetzer 1992). But Salmon also relinquished the conception of explanations as arguments, which circumvented the obligation to account for the function of the values of logical probabilities at all (Salmon 1971, pp. 70–71).

In his more recent work on causal-mechanistic explanations implementing conditions that are intended to restrict explanations to all and only causally-relevant factors, Salmon at least acknowledges the importance of probability values, not only in relation to the maximally homogeneous reference class to which the event to be explained should be assigned, but also for every other partition of the reference class (Samon 1984, pp. 36–37). His causal-mechanistic account thus appears to improve upon his statistical-relevance model, not only by appealing to causal-relevance relations in lieu of statistical-relevance relations but by emphasizing the importance of associated probability values.

In this respect, his approach less resembles the model advanced by Paul Humphreys than it does the model proposed by Peter Railton, both of which also dispense with degrees of nomic expectability. On Railton's D-N-P model, for example, complete probabilistic explanations assume the following form:

18a] (1) a theory entailing a law of form (2).
(2) (x)(Ax $\Rightarrow$ Probability Bx = r).
(3) Aa.

---

(4) Probability Ba = r.
(5) (Ba).

where the explanandum sentence appears as an addendum (Railton 1978, 1981). Railton asserts, "what occurs by chance must be explained by chance", but his model explains why an event had a certain propensity, not why it occurred.

On Humphreys' aleatory approach, explanations assume the canonical form, 18b] 'x because y, despite z', where 'y' specifies contributing factors and 'z' counteracting factors that made a difference to the occurrence of the event described by explanandum sentence x (Humphreys 1981, 1983, 1985). He contends (a) that probabilistic explanations do not require probability values, (b) that explanations can be adequate even when they are incomplete; and (c) that adequate explanations can do without subsumption under laws. Some of these are motivated by the mistaken belief that maximal specificity conditions require that an explanation of death from lung cancer, for example, would have to include every factor that made an historical contribution to that death, no matter how remote in space/time. But propensities satisfy Markov conditions that exclude all but the most contemporaneous factors: causal connectability does not guarantee explanatory relevance. Present effects have to be explained by present causes (Fetzer 1983).

Abductive Inference I

What is most important about these three accounts relative to the present inquiry, however, is that they supply no roles for degrees of nomic expectability. This is a crucial blunder, because these logical probabilities forge the linkage between explanation, prediction, and inference. These bracketed values not only represent estimates of the truth frequency with which explanandum sentences of that form may be expected to be true over extended sequences of trials, given the truth of premises with the form of those explanans sentences, but also functions as a designation of the degree of nomic expectability for such outcomes to occur on each single trial with the force of logical necessity, a benefit derived from a single-case account of probability, which in turn provides an ingredient that is essential to inference to the best explanation (Fetzer 1981).

When the likelihood L of hypothesis h, given evidence e, is defined as equal to the probability P of e, given h, this relationship may be formalized as follows:

19a] $L(h/e) = P(e/h)$

which might be applied to probability hypotheses generally. We need a special version that applies to inferences to laws in relation to the nomic expectability of evidence describing the initial conditions and the explanandum, specifically:

19b] $$NL(L/IC \,\&\, E) = NE(E/L \,\&\, IC)$$

which asserts that the nomic likelihood NL that law L applies, given the initial conditions IC and the explanandum E, equals the nomic expectability NE of the explanandum E, given an explanans of IC and the law L (Fetzer 1981, p. 223).

This conception permits the application of Hacking's "law of likelihood" as the fundamental principle of reasoning for inference to the best explanation, where evidence e supports hypothesis h1 better than it does hypothesis h2 when (1) the likelihood of h1, given e, exceeds that of h2, given e, and (2) the evidence e supports h1 better than it supports h2 when the likelihood ratio of h1 to h2 exceeds 1 (Hacking 1965, pp. 70–71). While satisfying these relations may make h1 preferable to h2, however, that does not mean h1 is therefore acceptable as well, which depends upon the quantity and quality of the available evidence e.

Abductivist Inference II

This conception of nomic likelihoods appears to apply to inferences to laws as explanations of relative frequencies as well as to explanations that use laws to draw inferences to explanations for singular events. Consider, to begin with, an explanation for the occurrence of decay by an atom of polonium 218, namely:

20] $$\begin{array}{ll} (x)(t)[Pxt^{218} \;=>\; (Txt = 3.05\,\text{min} = 1/2 \Rightarrow Dxt + 3.05\,\text{min})] & \\ Pat^{218} \;\&\; Txt = 3.05\,\text{min}. & \\ \hline Dat + 3.05\,\text{min}. & [1/2] \end{array}$$

which asserts that, if x were an atom of polonium 218, then subjecting x to a time trial of 3.05 min duration would bring about its decay with propensity 1/2, because of which the nomic expectability for that outcome is [1/2].

That the halflife of polonium 218 is 3.05 min not only means that a single atom has a propensity of 1/2 to decay during a 3.05 min interval but also implies that that same atom has a propensity of 1/2 to not undergo decay during that same interval. It also implies that, for large numbers of atoms of polonium 218 given at a specific time, very close to one-half will still exist 3.05 min later, the remainder having disintegrated by decay (cf. Hempel 1966, p. 66). If the halflife of polonium 218 were not know, repeated observations of decay on this order would support such an inference as the hypothesis that provides the best explanation for the frequency data.

Alternative hypotheses that might deserve consideration would include those that cluster around the observed relative frequency of decay in large samples, which would have values close to 3.05 min. Those hypotheses would have high likelihoods by virtue of making those outcomes highly probable, which indicates that more than one hypothesis can have high likelihood within the framework of this non-probabilistic account (Fetzer 1981, p. 276). When repeated sequences of random trials are conducted under suitable test conditions and yield stable relative frequencies for decay, the hypothesis with the highest likelihood would deserve to be accepted (Fetzer 1981, Chapter 9).

Indeed, measures that are internal to the evidence can provide a standard for determining that sufficient evidence has become available, namely: when the relative frequencies for possible outcomes have "settled down" and display stable values that have resisted our best efforts to change them, then we have appropriate evidence for the acceptance of the most likely hypothesis. While the severity of the tests that should be employed for this purpose tends to depend upon the seriousness of the consequences of making mistakes, objective criteria for acceptance have been advanced that make the sufficiency of evidence a function of the distribution of the data (Fetzer 1981, esp. pp. 244–254).

Abductivist Inference III

One such approach proposes the degree of divergence of the observed data from a normal distribution as a suitable measure of epistemic caution, where no hypothesis may be accepted from the set of alternatives unless its likelihood is greater than this degree of divergence, which might be measured by the Lévy distance (Gnedenko and Kolmogorov 1954, p. 33). Hypotheses with high likelihood on the data may be acceptable *even* when our measure of confidence happens to be low, while hypotheses with low likelihood on the data will be acceptable *only* when our measure of confidence happens to be high (Fetzer 1981).

Considerations of this kind appear to carry over to the explanation of single events, which may occur in ordinary contexts as well as in scientific inquiries. The explosion of TWA Flight 800 on 17 July 1996, for example, in which all of the 230 people on board were killed, involved a Boeing 747. At least three hypotheses have been advanced to account for this explosion, including (H1) that the catastrophe occurred as the result of a terrorist's bomb, (H2) that the plane was hit by an errant missile, and (H3) that an unsuspected design failure might have caused the fuel tank to blow up. The case has been difficult to resolve, not because of a lack of possible explanations, but because of an absence of evidence.

After expending enormous resources in the recovery and the reconstruction of the aircraft, however, the situation appears to have settled down. Initial reports ranged from eyewitness testimony to expert analyses that appeared in a wide range of publications, from tabloids, such as the *Star* (26 November 1996) in support of (H1) to national magazines, such as *Newsweek* (23 December 1996) in opposition to (H2), and our nation's newspaper of record, *The New York Times* (24 May 1997) in support of (H3). An extensive investigation by the FAA has now concluded that an unsuspected design failure appears to have been at fault.

Several elements have contributed to this inference, including simulations of the kind of explosions that might be brought about by such a flaw. Concern has now spread to other aircraft with similar designs, including, especially, the Boeing 737, among the most popular commercial aircraft in the world, with more than 1,100 in use today in the United States alone (Wald 1998). The discovery of worn wire insulation in most of 13 planes subject to preliminary inspection has increased concern. While the probability for the explosion would be high given either (H1) or (H2), the absence of evidence supporting them and the discovery of evidence corroborating (H3) has made (H3) an acceptable hypothesis.

Abductivist Inference IV

Events of this kind bring home the importance of the difference between relative frequencies and causal tendencies. Even if the odds of dying in an aircraft accident are remote, say, on the order of 1 in 250,000, that does not mean that the chances of survival for those who are actually involved in aircraft accidents are 249,999 to 1. On the contrary, although aircraft accidents are relatively rare (and therefore, with respect to causes of death in general, only about 1 in 250,000 are due to aircraft accidents), the propensities for death or injury for passengers involved in aircraft accidents are rather high.

By comparison, a superabundance of evidence was available in the murder of Nicole Brown and Ronald Goldman on 11 June 1994. Most men who abuse women do not murder them, but more than half the abused women who are killed were murdered by their abusers. If we consider our initial hypothesis in this case to be (H1) that O. J. Simpson committed the crime, then possible alternative explanations include (H2) that it was a drug-related hit, where Faye Resnick, a friend of Nicole, was the intended victim, and (H3) that the true target was Ronald Goldman, where Nicole was an innocent bystander.

Since Faye Resnick was a casual user of cocaine who had lots of money to support her habit, (H2) really will not do; and since almost no one outside of the restaurant where he worked would have known that Ron was returning Juditha Brown's glasses to her daughter, (H3) cannot be taken seriously. The evidence – including J. O.'s blood at the scene, in the Bronco, in his home, and so on, the matching gloves, the hair follicles, the Bruno Magli shoe prints, the photographs, etc. – all appeared to provide overwhelming evidence that (H1) had to be accepted as true. So how could the jurors have acquitted him?

The jurors in this case seem to have believed that if any part of the prosecution's case was open to doubt, then they could properly disregard it all. The jury forman, Armanda Cooley, for example, acknowledged that the jurors were unable to "explain away" the presence of O. J.'s blood at the scene (Cooley 1995, p. 202). Without some reasonable alternative explanation, however, the only reasonable explanation appears to be that the blood of the accused was at the scene of the crime because the accused was at the scene of the crime, which he committed. If it cannot be "explained away", then it has to be "explained" by hypothesis (H1), which implicates O. J. Simpson in the crime. That explains it.

Abductivist Inference V

Criminal cases are excellent examples of inference to the best explanation on several grounds, since they usually commence with puzzlement over who might have committed the crime and then engage in speculation (often using the heuristic criteria of motive, means, and opportunity), where a list of suspects represents the alternative hypotheses. As investigations proceed, some of the suspects may be eliminated because their commission of the offense is no longer consistent with the available evidence, while other suspects may become stronger. As sufficient evidence becomes available, the prosecutor may bring an indictment and attempt to establish a case beyond reasonable doubt.

The most fascinating illustration of inference to the best explanation that I have encountered within purely scientific contexts, however, has appeared in the field of cognitive ethology. Carolyn Ristau has extensively studied the behavior of the piping plover, which feigns broken-wing displays that distract predators when they threaten its nest and young. She describes the plover's full display as consisting of outstretched widely-arched wings that flutter and drag along the ground, offering persuasive evidence of injury, where the observer may trudge hundreds of meters after it only to see it agilely fly away at a point that is far removed from the nest and young (Ristau 1991, p. 94).

In order to evaluate the prospect that the piping plover is acting purposefully in displaying this behavior, Ristau introduces a set of alternative explanations, which include (H1) that this might be a reflex or fixed action response, (H2) that this might be a manifestation of conflicting motivation, (H3) that this might be brought about by approach-avoidance tendencies, (H4) that this might be an innate stimulus-response pattern of behavior, (H5) that this might be a behavior that the plover has acquired from experience, and (H6) that it might be purposeful behavior that is intended to secure a specific objective or goal.

By skillfull deducing consequences that would have to be true, given each of these hypotheses, and conducting systematic experiments that vary variables that would have to make a difference if various hypotheses were true, Ristau establishes that the plover's injury-feigning is non-random, does not simply lead away from its young or from the intruder, does not inconsistently lead away from its nest, is not rigid and inflexible, and is not acquired by repeated exposure to similar behavior, where the only alternative hypothesis that can explain the available evidence is that the plover's display is purposive behavior that is intended to secure the goal of protecting its nest and its young.

Abductivist Inference VI

Ristau's brilliant study, which reports data derived from 19 different experiments and from 10 birds that were members of 4 different pairs of piping plovers and 2 different pairs of Wilson's plovers, which yielded 45 displays of broken-wing behavior sufficiently detailed for analysis, could hardly offer a better illustration of inference to the best explanation. She takes great pains to make sure that every reasonable alternative hypothesis receives consideration, she considers the consequences that attend each of those alternatives, she conducts systematic experiments to test them, and, when sufficient evidence becomes available, she accepts the best explanation that is available.

However impressive this might initially appear, it may all be for naught if Peter Lipton is correct in suggesting that there is a fundamental difference between the *likeliest* explanation and the *loveliest* explanation (Lipton 1991, pp. 61–63). Thus, in Lipton's sense, the likeliest explanation is the best warranted explanation, which may be the most conservative relative to the available evidence by going least far beyond it, while the loveliest explanation is the most general explanation, which provides the greatest understanding. And Lipton is certainly correct in supposing that there are explanations of different kinds.

Some explanations appeal to premises that are analytic rather than synthetic and provide understanding rooted in language rather than in phenomena. That bachelors are unmarried and Donald is a bachelor (again), for example, supports the inference that Donald is unmarried (again), but it certainly does not appeal to a natural law. That individual things that are green look green when they are observed under standard conditions of lighting by persons of normal vision, moreover, may explain why they look green, but the degree of understanding it provides is restricted to the presence of properties. Cases of this kind also show how easily instantiations of causes without laws can occur.

Lipton's own example, namely: that smoking opium induces sleep because of its dormative (or sleep-inducing) powers, initially resembles the color case, but properly understood appears to attribute a permanent property to everything that is opium. It differs from accidental generalizations that attribute transient properties, such as that every Volkswagen is painted gray, which at one time was the case in America. Every VW was such that, if you looked at it under standard conditions of lighting and had normal vision, it looked gray. His example appears more like the lawlike sentence that emeralds are green, which implies that, when observed under suitable conditions, they look green.

Abductivist Inference VII

Even though there are kinds of explanations that do not appeal to lawlike premises and cannot qualify as scientific, which therefore do not provide the same kind of understanding as do scientific explanations, there is no apparent basis here for distinguishing between likely and lovely *scientific explanations*. Moreover, the conception of explanation that has been presented above seems to have other virtues that support precisely the conditions of adequacy thereby defined. In particular, the distinction between *possible* explanations and *adequate* explanations appears useful in separating science from non-science.

A debate over whether creationism should be taught as part of the science curriculum in public high schools, especially, has generated considerable controversy in the United States in recent times. Examples of classic creationist hypotheses include (CH-1) that God created the world and everything therein exactly as it exists today, (CH-2) that God created the world and living things in fixed and unchanging forms, and (CH-3) that God created the world and all living things using the causal mechanisms of evolution. Presumably, such hypotheses should be taught in the science curriculum only if they are scientific.

Using the conditions of adequacy (CA-1) through (CA-3), however, it can be shown that none of these hypotheses properly qualifies as a possible scientific explanation. (CH-1), for example, does not satisfy (CA-1), the derivability condition, because no specific explanandum follows from it. It is compatible with the world, no matter what, and thus possesses no empirical content. (CH-2) implies that all living things exist in forms that are fixed and unchanging and thus satisfies (CA-1), but it makes no appeal to any lawlike premise and therefore does not satisfy (CA-2). It specifies no lawful relationship between God and his creations, whether or not they exist in fixed and unchanging forms.

(CH-3) asserts that God made the world and all living things using the causal mechanisms of evolution. Assuming the causal mechanisms of evolution – including genetic mutation, natural selection, sexual reproduction, sexual selection, genetic drift, group selection, artificial selection, and genetic engineering – are lawful and sufficient to explain the origin of species, then this hypothesis satisfies (CA-2). But God becomes an explanatorily irrelevant factor, whose presence or absence makes no difference. It therefore cannot satisfy (CA-3). But if these hypotheses cannot satisfy the conditions for possible explanations, they are not scientific alternatives and should not be taught as if they were.

Abductive Inference VIII

Lipton also suggests that other explanations may be lovely without being likely, offering conspiracy theories as his illustration (Lipton 1991, pp. 61–62). He observes that conspiracy theories might acquire considerable explanatory power by showing that many seemingly unrelated events and apparent coincidents are actually related: "If only it were true, it would provide a very good explanation. That is, it is lovely. At the same time, such an explanation may be very unlikely, accepted only by those whose ability to weigh the evidence has been tilted by paranoia." But our normal access to truth is by way of inference.

Sometimes the evidence supports conspiracy theories. In 1992, I organized a research group in an attempt to place the investigation of the death of JFK on an objective and scientific foundation. The possible explanations include (H1) that he was killed by a lone, demented gunman, (H2) that the Mafia did it, (H3) that pro- or anti-Castro Cubans were responsible, (H4) that the KGB killed him, or (H5) that he died as the result of a *coup d'etat* involving the CIA, the Mafia, anti-Castro Cubans and powerful politicians, including LBJ and J. Edgar Hoover, with financing from Texas oil men and parts of the military-industrial complex.

My collaborators, who include a world authority on the human brain who is also an expert on wound ballistics, a Ph.D. in physics who is board certified in radiation oncology, and various experts on other aspects of the evidence in this case, have discovered that autopsy X-rays have been fabricated, in one instance, by concealing a massive blow-out to the back of his head, in another instance, by adding a 6.5 mm metal object; that diagrams and photographs of a brain in the National Archives are of the brain of someone other than JFK; and that the Zapruder film has been massively edited using highly sophisticated techniques.

These discoveries have substantial impact on alternative theories, since the Mob, for example, would not have had the reach to extend into Bethesda Naval Hospital to fabricate X-rays under control of the Secret Service and officers of the U.S. Navy; pro- or anti-Castro Cubans could not have substituted diagrams and photographs of the brain of someone else for that of JFK; and the KGB, even if it had the ability to do so, could not have gained possession of the Zapruder film in order to subject it to extensive editing. Nor could any of these things have been done by Lee Harvey Oswald, who was incarcerated or already dead.

These findings, which strongly suggest that the assassination was the result of a well-organized, high-level conspiracy involving officials of the government, has been corroborated by the discovery of more than 15 indications of Secret Service complicity, from the use of an improper motorcade route to vehicles out of sequence to the driver bringing the limousine to a halt after bullets began to be fired. Were we to select only eight or nine of these many events and treat them as simple coincidences that happen now and then, say, one time in ten, then their improbability of simultaneous occurrence would equal 1 in 100,000,000 to 1 in 1,000,000,000, suggesting they did not happen by chance (Fetzer 1998, 2000c).

The Justification of Induction

Likelihood measures thus work well in cases of various kinds. Even if evolution imparts only a low probability to the origin of life, if we assign creationist hypotheses zero or infinitesimal likelihoods to render them mathematically tractable, likelihood measures of support will overwhelmingly favor evolution. And if powerful conspirators with money and time could only be expected to successfully execute between 85% and 90% of their diabolical plans, the available relevant evidence concerning the assassination of JFK would still render the occurrence of a covert *coup* vastly more likely than any of its alternatives.

Scientific habits of mind and patterns of inference should be carried into the public sphere of daily life, where they can contribute immeasurably to the successful resolution of difficult problems. While abductivism seems to overcome the most important objections to its deductivist alternatives, especially by supplying solutions to problems of acceptance, the advantages of deductivism over inductivism are profound. Popperian conceptions of laws of nature and of probabilities as propensities appear to be indispensable to an adequate account of science. Corroboration is vastly superior to confirmation.

While inductivism can account for inferences from samples to populations and deductivism can account for inferences from the observed to the unobserved, only abductivism can account for inferences from the past to the future. The pragmatic vindication of the straight rule cannot be sustained, because there is no reason to believe that the world's history will go on forever. And, even if it did, we could never know that the relative frequencies we observe are even roughly close to the values of the limits we infer. The solution to the problem of induction must be based upon an ontic conception of randomness.

What the single-case propensity account ultimately supplies is an account of natural laws that applies whether the history of the world is short or long. Even if the world *is* as we believe it to be with respect to the laws of nature, it remains logically possible that the future might not resemble the past in those very respects; but *if* the world is as we believe it to be with regard to natural laws, it is not physically possible that the world might not resemble the past in those same respects, which appears to be the strongest solution to the problem of induction that empirical procedures are able to provide.

References

Cooley A et al (1995) Madam Foreman. Dove Books, Beverly Hills, CA
Fetzer JH (1977) A world of dispositions. Synthese 34(1974):397–421
Fetzer JH (1981) Scientific knowledge. D. Reidel, Dordrecht/Holland
Fetzer JH (1983) Probabilistic explanations. In: Asquith P, Nickles T (eds) PSA 1982, vol 2. Philosophy of Science Association, East Lansing, MI, pp. 194–207
Fetzer JH (1992) What's wrong with Salmon's history. Philos Sci 59(June 1992):246–262
Fetzer JH (1993) Philosophy of science. Paragon House, New York
Fetzer JH (ed) (1998) Assassination science. Open Court, Chicago, IL
Fetzer JH (2000a) Editor's introduction. In Fetzer JH (ed) Science, explanation, and rationality: aspects of the philosophy of Carl G. Hempel. Oxford University Press, New York, pp xv–xxix
Fetzer JH (2000b) The paradoxes of Hempelian explanation. In Fetzer JH (ed) Science, explanation, and rationality: aspects of the philosophy of Carl G. Hempel. Oxford University Press, New York, pp 111–137
Fetzer JH (ed) (2000c) Murder in Dealey plaza. Open Court, Chicago, IL
Gnedenko BV, Kolmogorov AN (1954) Limit distributions for sums of independent random variables. Addison-Wesley, Reading, MA
Hacking I (1965) Logic of statistical inference. Cambridge University Press, Cambridge
Hartshorne C, Weiss P (eds) (1935) The collected papers of Charles Sanders Peirce, vol 5. Harvard University Press, Cambridge, MA
Hempel CG (1962) Deductive-nomological vs. statistical explanation. In: Feigl H, Maxwell G (eds) Minnesota studies in the philosophy of science, vol III. University of Minnesota Press, Minneapolis, MN, pp 98–169
Hempel CG (1965) Aspects of scientific explanation. The Free Press, New York
Hempel CG (1966) Philosophy of natural science. Prentice-Hall, Englewood Cliffs, NJ
Hempel CG (1968) Maximal specificity and lawlikeness in probabilistic explanation. Philos Sci 35(2):116–133

Humphreys P (1981) Aleatory explanations. Synthese 48(2):225–232
Humphreys P (1983) Aleatory explanations expanded. In: Asquith P, Nickles T (eds) PSA 1982, vol 2. Philosophy of Science Association, East Lansing, MI, pp 208–223
Humphreys P (1985) The chances of explanation. Princeton University Press, Princeton, NJ
Jepson A (1998) I was squished by a steamroller. The National Enquirer (19 May 1998), p 10
Lipton P (1991) Inference to the best explanation. Routledge, London
Popper KR (1959) The propensity interpretation of probability. Br J Philos Sci 10:25–42
Popper KR (1965) The logic of scientific discovery. Harper & Row, New York
Popper KR (1990) A world of propensities. Thoemmes Anti-quarian Books, London
Railton P (1978) A deductive-nomological model of probabilistic explanation. Philos Sci 45: 206–226
Railton P (1981) Probability, explanation, and information. Synthese 45:233–356
Reichenbach H (1937) Experience and prediction. University of Chicago Press, Chicago, IL
Reichenbach H (1949) The theory of probability. University of California Press, Berkeley, CA
Ristau C (1991) Aspects of the cognitive ethology of an injury-Feigning bird. In Ristau C (ed) Cognitive ethology. Lawrence Erlbaum, Hillsdale, NJ, pp 91–126
Ruben S (1985) Handbook of the elements. Open Court, La Salle, IL
Salmon WC (1967) Foundations of scientific inference. University of Pittsburgh Press, Pittsburgh, PA
Salmon WC (1971) Statistical explanation and statistical relevance. University of Pittsburgh Press, Pittsburgh, PA
Salmon WC (1984) Scientific explanation and the causal structure of the world. Princeton University Press, Princeton, NJ
Wald M (1998) Agency grounds scores of 737's to check wiring. The New York Times (11 May 1998), p A1, A 14

Index

Q

R

S