SYNTHESE LIBRARY

STUDIES IN EPISTEMOLOGY,

LOGIC, METHODOLOGY, AND PHILOSOPHY OF SCIENCE

VOLUME 226

PATRICK SUPPES

PATRICK SUPPES
Lucie Stern Professor of Philosophy, Stanford University

MODELS AND METHODS IN THE PHILOSOPHY OF SCIENCE: SELECTED ESSAYS

SPRINGER-SCIENCE+BUSINESS MEDIA, B.V.

Library of Congress Cataloging-in-Publication Data

Suppes, Patrick, 1922-
Models and methods in the philosophy of science : selected essays / by Patrick Suppes.
p. cm. -- (Synthese library ; v. 226)
Includes index.
ISBN 978-90-481-4257-6 ISBN 978-94-017-2300-8 (eBook)
DOI 10.1007/978-94-017-2300-8
1. Science--Philosophy. 2. Science--Methodology. 3. Physics. 4. Psychology. I. Title. II. Series.
Q175.S9394 1993
501--dc20 93-14714

ISBN 978-90-481-4257-6

Printed on acid-free paper

CONTENTS

PREFACE

The thirty-one papers collected in this volume represent most of the articles that I have published in the philosophy of science and related foundational areas of science since 1970. The present volume is a natural successor to *Studies in the Methodology and Foundations of Science*, a collection of my articles published in 1969 by Reidel (now a part of Kluwer).

The articles are arranged under five main headings. Part I contains six articles on general methodology. The topics range from formal methods to the plurality of science. Part II contains six articles on causality and explanation. The emphasis is almost entirely on probabilistic approaches. Part III contains six articles on probability and measurement. The importance of representation theorems for both probability and measurement is stressed. Part IV contains five articles on the foundations of physics. The first three articles are concerned with action at a distance and space and time, the last two with quantum mechanics. Part V contains eight articles on the foundations of psychology. This is the longest part and the articles reflect my continuing strong interest in the nature of learning and perception. Within each part the articles are arranged chronologically. I turn now to a more detailed overview of the content.

The first article of Part I concerns the role of formal methods in the philosophy of science. Here I discuss what is the new role for formal methods now that the imperialism of logical positivism has disappeared. The new imperialism of historicism is also now showing signs of fading. We have, I hope, entered a pluralistic era of irenic appreciation of many different ways of looking at science. In a closely related vein, Article 2 expresses skepticism about the methods used as yet to study the nature of scientific revolutions. I contrast the literature on these matters with methodologically more sophisticated approaches dealing with other subjects in history, especially economic history. Article 3 examines the limitations of the axiomatic method in ancient Greek mathematical sciences, which have too often been viewed themselves as paradigms of the axiomatic method. There is, I would claim, not too great a difference between the situation then and now regarding the use of axiomatic methods in science. Article 4 expresses skepticism about the unity of science and affirms the evident plurality of modern science, which has increased even more since this article was written. Article 5 contrasts the role of heuristics and the role of the axiomatic method in science and mathematics. Greater attention to working heuristics in various parts of science is something I hope to devote more effort to in the future. What is said here is a beginning. The final article (#6) in this section is on representation theory and the analysis of structure, a favorite topic of mine, since I first wrote about such matters in detail in the last chapter of my *Introduction to Logic*, published in 1957.

Part II is focused on causality and explanation. Much of what I have to say in the six articles in this part develops questions unanswered in my monograph *A Probabilistic Theory of Causality*, published in 1972. Article 7 deals with causal analysis of hidden variables with, as would be expected, special reference to quantum mechanics. This article could also easily have been placed in Part IV on the foundations of physics. Article 8 is a long reply to criticisms of some of my views on causality by the late Richard Martin. In answer to Martin's criticisms I appropriately modified some of my more sweeping claims and concentrated on the use of causal concepts in science. In Article 9 I deal with a phenomenon that I would have been skeptical of at one time, but no longer am, namely, giving good scientific explanations of unpredictable phenomena. With a modern emphasis on chaos it is a topic we shall be hearing a great deal more about in the future. In Article 10 dealing with conflicting intuitions about causality, I examine a number of puzzles, including Simpson's paradox, which as a Bayesian I do not find fundamentally paradoxical. Article 11 is the only article written jointly with another person, in this case Mario Zanotti, and is on probabilistic explanations. Here we prove the somewhat surprising theorem that deterministic hidden variables can be found if and only if the phenomenological variables have a joint probability distribution. In the last article of this part (#12) I deal with non-Markovian causality and show what conditions are required for non-Markovian causes to be transitive, a puzzle that was left open in my 1971 monograph. I use some earlier work of Eells and Sobers showing that a Markovian condition is sufficient for transitivity.

Part III concerns the foundations of probability and the foundations of measurement, two subjects that in my own work have been much intertwined. Article 13 gives a simple presentation of measurement structures of a variety of kinds. The elementary character comes from assuming finiteness of the basic domain and equal-interval placement of the objects, as would be characteristic of the fundamental theory of various measurement scales. Article 14 uses some of the same results to give a more general theory of the measurement of belief, with the use of upper and lower probabilities to characterize the nature of partial beliefs. Article 15 widens the domain of analysis to the logic of clinical judgment with emphasis on Bayesian and other approaches. This is the only article I have written on clinical judgment in medicine, although it is a subject I have been interested in for a long time, and have worked with a number of students on. I regret not having as yet written more in this direction, for it is a wonderful area in which to test one's intuitions about the usefulness of probability and also fundamental measurement. Article 16 gives arguments for randomizing, a topic of special importance to a sometime Bayesian like myself. As is clear from the article, I reject an extreme Bayesian viewpoint that sees no need for randomizing at all. I continue to be reasonably satisfied with the arguments given in this article, but I also see the need to accompany it with a more satisfactory technical discussion of randomization in finite sequences. As evidence that I am not completely a Bayesian about probability, Article 17 is concerned with propensity representations of probability. Here I am especially concerned to examine the way in which such representations of probability naturally arise objectively in physics. Two important cases are considered. One exemplifies Poincaré's method of arbitrary functions, and the other the important results one can get from the classical three-body problem to show that randomness can be found in as strong a form as you desire in simple deterministic systems consisting of a small number of particles. The last article in this part (#18) concerns general philosophical arguments about the choice between indeterminism or instability and the question of whether it matters which we choose. The concept of instability as a replacement for that of indeterminism has not been considered as much as it should in the philosophy of science. What I have to say needs much further development.

Part IV is concentrated on the foundations of physics. Article 19 is the only one written before 1970. It is an early article of mine from 1954 on Descartes and the problem of action at a distance, originally published in the *Journal of the History of Ideas*. In fact, it is a rewrite of a chapter in my 1950 doctoral dissertation on the problem of action at a distance. Article 20 concerns some open problems in the philosophy of space and time. Most of the problems seem to still be with us, even though the article was written 20 years ago. In a similar vein, Article 21 concerns Aristotle's concept of matter and it's relation to modern concepts of matter. Along the way I look at the theory of matter advocated by Descartes, Boscovich, and Kant. In spite of my long-term interest in Kant's philosophy of science, this is the only article in the volume dealing with Kant, and even the discussion here is somewhat *en passant*. Article 22 deals with Popper's analysis of probability and quantum mechanics. It is closely related to my views on the foundations of probability and could just as well have been placed in Part III. Article 23 represents my most recent views on the probabilistic foundations of quantum mechanics. It was an article written for a symposium for the statistician Jack Good, on the occasion of his 70th birthday. The theory of probability in quantum mechanics is a weak theory of the mean, as I try to explain in detail in the article. In my judgment the recognition of the probabilities being only mean distributions is a much more important fact about probability in quantum mechanics than is an attempt to develop a theory of probability defined on structures that generalize classical Boolean algebra.

Part V is concerned with the foundations of psychology. Article 24 is one I wrote in 1975 to chart a course from behaviorism to neobehaviorism. By "neobehaviorism" I meant the still standard practice in cognitive psychology to use as data psychological responses, but at the same time to admit the necessity of rich internal mental structures. Article 25 was written for a conference on structural models of thinking and learning and is concerned with learning theory for probabilistic automata and register machines, with applications to education research, in particular the learning of elementary mathematics. The work here amplifies and makes more concrete results that were part of my earlier work on stimulus-response theory of finite automata. Article 26 moves to perception and analyzes from both a historical and conceptual standpoint the question of whether or not visual space is Euclidean. I will not state the answer here; you have to read the article to find out. Article 27 is one concerned with Donald Davidson's views on psychology as a science. Davidson was a colleague of mine at Stanford in the 50's and we wrote several articles and a book together on decision making. It was a pleasure to write an analysis of the views on psychology he has expressed in articles published over the last several decades. Article 28 is a rather long one on current directions in mathematical learning theory. The first part surveys many different kinds of work including perceptrons and cellular automata, and the second part is devoted to amplifying and extending in a technical way my earlier work on stimulus-response theory of finite automata. Article 29 is on deriving models in the social sciences. Here I try to put emphasis on a theme I have emphasized elsewhere, but nowhere else in this volume really, namely, on the importance of studying and thoroughly understanding the methods used for deriving models, starting of course with the classical methods of deriving differential equations in physics. I try to draw some contrasts with various methods currently widely used in the social sciences. I end with a model aimed at very specific psychological ideas but which depends on classical methods for deriving the governing differential equation. Article 30 in this section returns to the theme of perception and analyzes the principle of invariance with special reference to perception. The first part deals with geometrical semantics for spatial prepositions. Here I extend some earlier work with Colleen Crangle. In the second part I return to the question of visual space being Euclidean and present more details of recent experiments and relevent theory. Anyone who is not persuaded by the answer given in Article 26, should also read the second half of this article to be fully convinced of what is the correct answer to the question of whether or not visual space is Euclidean. The final article (#31) is one that I wrote for a recent symposium on reductionism in science. I ask the question, Can psychological software be reduced to physiological hardware? In this case I will give away the answer. The article consists of four arguments for answering in the negative.

Broadly speaking, all of the articles on the foundations of psychology are concerned either with learning or perception. Some of the work in the part on probability and measurement could also easily be classified as belonging to the foundations of psychology. In any case, my most recent work in learning is not reflected in these articles for I am now concentrating above all on machine learning of natural language. This work is represented in the volume of my papers on language recently published, *Language for Humans and Robots* (1991b).

Among philosophers of science I am probably best known for advocating set-theoretical models and methods in studying particular problems. That interest and viewpoint are certainly reflected in the present volume. On the other hand, I like to stress that the strong empirical aspect of my work in the philosophy of science and also in science is present in this volume, although not as adequately represented. The general point I want to emphasize by mentioning this continuing concern with detailed empirical data is that I very much believe in a pluralistic approach to both philosophy and science. I am not at all dedicated to reducing all questions to those that can be framed in an explicit way within an appropriate set-theoretical model.

In preparing these thirty-one articles for publication, I have made no essential changes, but have of course corrected obvious mistakes or misprints. In addition I have tried to standardize the notation in a given area. However, given the diversity of topics covered and the conventions of notation reigning in different disciplines, it has not been feasible to introduce a completely standard notation throughout the volume. I have also deleted from some articles preliminary material of a standard formal kind which appeared in an earlier article, in order to avoid the most egregious forms of repetition. There is still some repetition in articles on common topics since it would have been awkward to delete all areas of overlap. I have also standardized the format of section headings, for the original articles were published in journals with many different styles. The references to the literature given in the various articles are all collected together at the end in a single list. Footnotes in the original articles are numbered beginning anew with each article. An index of authors referred to is given. In place of a subject index, there is a detailed table of contents at the beginning of the volume.

Acknowledgments for permission to reproduce the various articles are given at the bottom of the first page of each article, but thanks are extended here to the many editors and publishers who generously agreed to publication. Finally, I want to acknowledge the extensive work of Laura Tickle, sometimes with Emma Pease's assistance, in preparing this volume for publication in the now increasingly standard format of LaTeX. I also want to express my thanks to Kaija Lewis for her careful reading of the proofs and final LaTeX editing.

PATRICK SUPPES

Stanford, California

PART I

GENERAL METHODOLOGY

1

THE ROLE OF FORMAL METHODS IN THE PHILOSOPHY OF SCIENCE

1. THE END OF IMPERIALISM

In the period that ran from Frege to the Vienna Circle and Carnap, a strongly reductionist view of the philosophy of science held sway. The significant problems should be reducible to problems that could be formalized within logic. Those that could not be treated in this fashion should be dismissed as being too vague to be of interest. This description is something of a caricature but I shall not convert it into a genuine historical account. It is too familiar to all of us to be recounted here. My point is, rather, to emphasize that such a reductionist view of the place of formal methods in the philosophy of science is now faded. If anything, we face currently a new imperialism of historical methods, but I am doubtful that we will move to anything like a reductionist Hegelian position that all questions are ultimately historical in nature.

The present pluralistic and schematic view of the philosophy of science does have the danger of a lack of intellectual discipline. It can too easily be said that any sort of method is appropriate, but in most areas of

*Reprinted from *Current research in the philosophy of science*, P. D. Asquith and H. E. Kyburg, Jr. (Eds), 1979, pp. 16–27. East Lansing, MI: Philosophy of Science Association.

science, as well as in the philosophy of science, no broad fundamental theory seems achievable. We shall be faced for the foreseeable future with a plurality of problems and methods. Yet I am stating the thesis in a weaker form than I am prepared to affirm it. The absence of fundamental theory dominating a given area of science or the philosophy of science is a healthy and normal state of affairs. It is only during certain periods of aberration that we seem to have a fundamental theory that is at all close to being satisfactory in relation to the problems and data confronting us. Ptolemy had such a theory when he was flourishing in Alexandria and it was the case when Kepler rewrote the fundamental assumptions of astronomy and later when Newton rewrote them once again. Other examples of the hegemony of single fundamental theories can be drawn, given certain periods of chemistry, perhaps recently from certain parts of molecular biology, and even probably for a while in economics. It is my view, however, that since World War II the engines of empiricism have vastly outrun the horse-drawn carriages of theory. The facts that have been accumulated have simply overwhelmed theory in almost every area. The range of problems that have been posed has exceeded the capacity of theory to handle, and we are now in most scientific domains in a happy state of schematic and pluralistic approaches to most problems. High-energy physicists still like to announce that with just another order of magnitude of increase in the energies available we shall finally get to the ultimate simples of the universe. Most outsiders who have followed such repeated claims over the past two decades can scarcely be anything but skeptical, and marvel at this latest expression of philosophical naiveté.

It is also worth mentioning that at the very time historical methods are becoming increasingly important in the philosophy of science, formal methods are assuming a comparable importance in history. Some accessible examples of elementary quantitative research in history are to be found in the volume edited by Aydelotte, Bogue, and Fogel (1971). Technically more sophisticated instances are abundant in the restricted area of economic history.

Perhaps one of the best examples of the decline of theoretical hegemony is in psychology. During the 1940s and much of the 1950s, behaviorism was the dominant theoretical viewpoint and the organizing force, from a methodological standpoint, throughout the parts of psychology considered fundamental or basic by a large number of American psychologists. (The situation was rather different in Europe, but experimental psychology is the most American of all of the fundamental scientific disciplines, and so I shall not try to comment on the European scene.) The symbolic end of this hegemony was Chomsky's famous review (1959) of Skinner's book on verbal behavior (1977), but the thrust of behaviorism continued into the middle 1960s, and it is only in the present decade that the deep-lying nature of the theoretical disarray in psychology has become so apparent. It is my conjecture, along the lines of what I have already said in general, that the many separate theoretical enterprises now flourishing in psychology will not be replaced in the future by a single unifying discipline. In retrospect it is apparent that the claims of theoretical psychology, as exemplified for instance in Clark Hull's work of the 1930s and 1940s, are as intellectually absurd as the claims of Kant to establish an a priori foundation of natural science—indeed, most philosophers would probably find Kant more sensible, but then I think that is because they do not often look at his detailed arguments as, for example, in the *Metaphysical Foundations of Natural Science* but at the more general and therefore less absurd views to be found in the *Critique of Pure Reason.*

One of the points that I want to make in these remarks about the pluralism of theories is that, just as physicists have in the past been dominated by the search for ultimate simples and ultimate theory, so philosophers of science have sought certainty and completeness of theoretical foundations for science. This search runs all the way from Aristotle's views on demonstration in the *Posterior Analytics* to Carnap on the logical structure of the world. The decline of this long search for bedrock does not mean the end of the relevance of formal methods in the philosophy of science but rather the beginning of a new era of realism about their limitations as well as their potential.

The rest of this paper is concerned to expand on this last point. In the next section I discuss the variety of formal methods that seem appropriate in the philosophy of science. The following section is concerned with a survey of some of the open problems in the philosophy of science, in the analysis of which formal methods can play a role.

2. VARIETY OF FORMAL METHODS

The theme of my remarks is, as before, pluralistic in nature. I begin with the point that there is no agreed upon formal methodology to be used in the philosophy of science; a variety of methods are available and appropriate. It is no longer a philosophically interesting question to seek a single methodology.

At least four methods have a certain saliency; two of them have been prominent in the last half century. The four I have in mind are: formalization in first-order or second-order logic (extensional or intensional), formalization within set theory, the procedural approach characteristic of computer science, and the approach of informal rigor vividly supported by Georg Kreisel on various occasions. Broadly speaking, all of these methods are characterized by some form of mathematical approach to problems in the philosophy of science. It is not my thesis to argue that all problems can be brought within the framework of one of these formal approaches, but I would strongly resist the view that most problems of interest lie outside of such methods. There are, of course some philosophers who now have such a strongly historical orientation toward problems in the foundations of science that there is skepticism about the use of formal methods on any problems of significance. In my view, this is a momentary fashion that is mistaken. A properly balanced philosophy of science will encompass both formal and historical methods, and, indeed, some of the more sophisticated problems in the history of science can well be approached from a formal standpoint. Some of the claims about scientific revolutions, for example, would seem to require quantitative and statistical analyses of data if they are to be taken as serious claims having the same status as other scientific claims about natural or social phenomena.

Of the four methods of formalization I mention, logical formalization is certainly the one that has received the most attention from philosophers, and those who are not very conversant with science often tend to think of this as the only kind of formalization. Many interesting results have been achieved by such methods. Perhaps even more important, the widespread and almost universal familiarity among philosophers of science with the concepts of elementary formal logic have provided a useful common framework for discussion of a great variety of problems.

On the other hand, I have emphasized in numerous publications for many years the limitations of such formalization because of the richness of structure characteristic of most developed scientific theories. I have baptized this attitude of mine "to axiomatize a scientific theory is to define a set-theoretical predicate." I continue to think that such set-theoretical methods are appropriate for a wide variety of problems in the philosophy of science, and I have tried in various papers to make this view a concrete one by providing a number of examples. Such set-theoretical methods are not as widely used in the philosophy of science as methods of formal logic but they are certainly better known and more widely accepted than the last two methods I mentioned.

The procedural approach characteristic of computer science is just gaining currency in the philosophy of science. It is almost certainly the case that the pursuit of procedural or computational methods will be of considerably more importance than the further extension of set-theoretical methods in the development of fundamental psychological theories of cognition, learning, and perception. More generally, procedural approaches will probably come to be of greater importance theoretically in the philosophy of the social sciences than they now are. Because of the extensive use of computational methods by most empirically oriented social scientists, it seems likely that many future theoretical developments will depend upon the use of computational ideas, rather than set-theoretical concepts, for their formalization. Of course, such formalizations outside of set theory are already familiar in constructive parts of the foundations of mathematics but their use in science, as opposed to mathematics, will have a different flavor because of the extensive computer orientation of the methods.

I also want to mention the conjecture that procedural methods will turn out to be especially important in developing an appropriate theory of meaning and of comprehension for natural language. An example that would not be accepted by many people is the proposal that the meaning of a proper name may be taken to be the set of internal procedures by which the individual that uses or recognizes the proper name attaches properties or relations to the object denoted by the proper name. These procedures or programs internal to a particular language user are private and in detailed respects idiosyncratic. The appropriate notion for a public theory of meaning is a notion of equivalence or congruence of programs or procedures that is considerably weaker than this very strong sense of idiosyncratic individual program. If this viewpoint is at all correct, the search for any hard and fast sense of identity of meaning is mistaken—it is hidden away in the internal programming of each individual and is a notion of limited scientific interest. What we are after are congruences of procedures that can collapse these private features across language users to provide a public and stable notion of meaning.

Put still another way, procedural approaches are a natural method for incorporating intensional and constructive ideas within the same conceptual framework. For this reason especially they would seem to have a very considerable future as an appropriate method of formalization in the philosophy of science.

Kreisel's lively defense of informal rigor (see especially his 1967 article) has been directed at overly formalistic and positivistic conceptions of the foundations of mathematics and rather little at the philosophy of science. His views are a proper propaedeutic to those who have been too enthusiastic about formalism and not sufficiently attentive to the need for informal and intuitive ideas of a definite nature about a subject in order to have significant ideas about it. In my own view, some of the analyses offered of the notion of causality in the philosophy of science suffer from a lack of informal rigor, that is, from a lack of serious attention to detailed scientific examples and the test of the formal ideas proposed against a variety of systematic intuitive results. Kreisel has made the point to me in several conversations that the use of set-theoretical methods in the philosophy of science is actually an example of informal rigor because it is the intuitive notion of set that is being used and not one axiomatized within first-order logic.

3. VARIETY OF OPEN PROBLEMS

I have divided this discussion of open problems to which formal methods are relevant into three parts, one dealing with theories, one with methodology, and one with problems of experimental evidence.

Theories. The historical development of physics is being investigated in fascinating ways by historians of science and by philosophers of science using an historical approach. All of us will learn a great deal from this work, whether we concentrate on Neugebauer's history of ancient astronomy (1975), current work concerned with the history of quantum mechanics, or any period in between. Although I am a strong advocate of formal methods, I am also an inveterate reader who has learned much from a variety of historically oriented works. My own generally skeptical views about having certain knowledge of a clear and definite sort about any complex phenomena have been much reenforced by Neugebauer's skepticism concerning the possibility of ever tracing causal influences in the history of science.

But historical study of fundamental scientific disciplines is, I want to insist once again, not the whole story for the philosophy of science. There are many questions of great philosophical interest that are no more historical than the corresponding development of new science. All such work, of course, should be properly historical in paying attention in technical detail to prior work that is relevant. But this is not what is meant by historically oriented studies and I only mention it because there has been a tendency in the philosophy of science to write about subjects without prior attention to the serious previous work. One of my own favorite examples of the view that ignorance is best is Norman Campbell's work on the theory of measurement, which reflects no serious acquaintance with the deeper and more sophisticated earlier work of Helmholtz and Hölder.

In those domains of physics that are properly regarded as being of fundamental philosophical interest, the number of formal problems of foundational interest is too large to enumerate here. It is easy to give a long list of open formal problems in quantum mechanics alone, the most important empirical scientific theory of this century. We are as yet far from understanding the role of probability in quantum mechanics. At an even more general level, there is still dispute about whether the final formulation of quantum mechanics should depend upon a nonstandard special quantum logic. The role of the theory of measurement and the explicit theory of the observer is also still a subject of controversy and one whose clarification bears on a number of significant problems of theory construction.

It is sometimes thought that the use of formal or axiomatic methods in the study of such problems is quite foreign to the work of physicists themselves and should be regarded as a new kind of scholasticism introduced by philosophers seeking to impale new angels on new needles.

Without entering into a dialectical discussion of this matter, I want to quote one significant piece of evidence to the contrary, the introductory paragraph of the well-known book on axiomatic quantum field theory by Bogolubov, Logunov, and Todorov (1975) on the place of the axiomatic approach in physics.

> It is widely believed that axiomatization is a kind of polishing, which is applied to an area of science after it has been, for all practical purposes, completed. This is not true, even in pure mathematics. Admittedly, the modern axiomatization of arithmetic and Euclidean geometry marked the completion of these disciplines (although at the same time it stimulated a new science—mathematical logic, or metamathematics). For most areas of contemporary mathematics, however, such as functional analysis, axiomatization is a fundamental method of exploration, a starting point. (Of course, the system of axioms may be modified as the subject develops.) In theoretical physics, since the time of Newton, the axiomatic method has served not only for the systematization of results previously obtained, but also in the discovery of new results. (1975, p. 1)

I mention as examples two other areas of science in which open problems exist for which formal methods are appropriate, and both the analysis and results would be of philosophical interest.

One concerns the notion of causality in the work of modern mathematical economists and econometricians. There is an increasingly technical literature on the use of causal notions in economics, especially as intertwined with a variety of detailed statistical methods for the analysis of economic data. To some extent the methods are mathematically intricate and relatively sophisticated because of the absence of the possibility of experimentation in economics. More powerful analytical methods are required in order to make a firm identification of causal phenomena. As far as I know there has not been any really thorough formal analysis from the standpoint of the philosophy of science of this large economic literature (a beginning may be found in Suppes (1970)). It is my belief that we as philosophers could learn much from this literature and at the same time we could bring to it a philosophical perspective that could contribute something as well. A simple but elegant technical example of this literature is to be found in Hosoya's (1977) proof of the general equivalence of the Granger (1969) condition for noncausality and the Sims (1972) condition. It is my impression that economics is almost the only science at present in which one can find unabashed technical discussions of causality in general terms and with careful theoretical development of concepts. (I exclude in this remark the use of causality principles in physics which refer essentially only to precedence in time.)

As a second example I mention the classical mind-body problem in the new guise of software and its independence of hardware. One need not go far in current neuroscience to realize that we are probably further from understanding how the mind works at a neural level than we ever thought we would be, at this point in time, say 30 or 40 years ago. The more work that is done, the greater the mystery deepens and it now seems an appropriate formal problem for psychology to establish under the weakest and most reasonable assumptions possible the impossibility of reducing complex phenomena to neurophysiological phenomena. The analogy here is that from an inspection of computers one can say very little about the kind of software that will be written for them, and if the problem were approached, once the program is encoded, in a purely physical fashion, I do not doubt that we would be unable ever to discover what the program is. There is even some skepticism that a large operating system encoded in binary digits could be fully understood if no external cognitive guides were provided. In any case, the relation between brain and mind is much worse because we do know so little about the detailed physical basis for encoding complex mental events. It would be interesting to formulate a variety of theorems about the impossibility of a reduction of mind to brain. Such theorems should play the same role in psychology that impossibility theorems about hidden variables play in quantum mechanics. The rigorous pursuit of the details should prove enlightening and should give us new ideas about how to formulate the concepts of psychology in a way that is properly independent of neurophysiology. (A somewhat more detailed discussion of these matters is to be found in Suppes (1975).)

Methodologies. The two decades running from 1945 to 1965 were marked by a more rapid expansion of work in mathematical statistics than in any other period of history. This was the time when decision-theoretic ideas were made the center of much of the theoretical literature in statistics.

The spread of applied statistics in the empirical sciences has also been more marked in the period since World War II than at any other time. At least until rather recently, these developments in mathematical statistics at both the theoretical and applied level have been largely ignored by philosophers, even those interested in the foundations of probability and induction.

Part of the reason for this separation between the statistical and philosophical literature on the foundations of induction has been the development of a separate strand of work by philosophers, generally labelled *confirmation theory.* A characteristic feature of confirmation theory has been its use of particular formal methods, primarily those of elementary logic. As a consequence, it has been difficult to make contact with the more elaborate and mathematically more technical machinery of mathematical statistics. Recently this situation has begun to change and there is now an increasing number of philosophers becoming knowledgeable about the foundations of statistics.

I would like briefly to mention some of the problems of statistical methodology of great importance in applied work which have not yet received definitive solutions. These are problems of obvious philosophical interest, and they illustrate the important role of formal methods in the foundations of probability and induction. First, there has been in the last decade and a half a new and extensive body of work on the concept of randomness. In the hands of Martin Löf and others, this concept has now begun to have implications for the theory of statistics as well as probability. A related question is the Bayesian problem of justifying random-sampling procedures. Formal and axiomatic analysis of the basis for random sampling continues to be a prominent problem and one that deserves philosophical consideration. More generally, the theory of finite samples and the admissibility or inadmissibility of the concept of an infinite population from which samples are drawn need further analysis. Certainly finitistic Bayesians would not be willing to admit the appropriateness of the concept of an infinite population, but the concept, on the other hand, has a long history of use and development in objective statistical theory.

The theory of experimental design has had intensive technical development since World War II, but the foundational principles are still in an unsatisfactory state. The recent flurry of interest in Bayesian statistics has not yet produced a satisfactory Bayesian theory of experimental design. The foundational literature that derives from de Finetti has as yet scarcely made contact with the technical problems of design. Almost certainly the principle of exchangeability, whose importance de Finetti has emphasized since the late 1920s, should play a prominent role in the foundations of the relevant Bayesian concepts, but much remains to be done to work out the formal theory.

This last mention of the role of the principle of exchangeability provides an opportunity to make a point that is implicit in what I have already said. I see no basis for drawing a sharp separation between the work that is to be done by philosophers interested in these problems and by statisticians with similar interests. There is a necessary and even a desirable overlap. It does not mean that philosophers must become professional mathematical statisticians in order to pursue problems of the foundations of statistics but it does mean, as in the case of the philosophy of science in other areas, as, for example, the philosophy of quantum mechanics, that detailed knowledge of relevant scientific work is a necessary background of informed philosophical analysis. What I would urge as strongly as possible is that philosophers meld their own formal methods with those of statisticians, in order to concentrate on the conceptual problems of interest. In this necessarily superficial survey of problems, I have especially mentioned problems of randomness of sampling, and of experimental design, because it is just these problems of central importance in the conceptual foundations of statistics that have received quite inadequate analysis in the philosophical literature on probability and induction.

Problems of experimental evidence. I have for a long time worried about formal models of data (1962). It is still my conviction that the philosophy of science has a significant contribution to make to the theory of how experimental evidence is used to test or evaluate scientific theories. Obviously a variety of problems that arise in this context are primarily statistical in character and would properly be treated under the heading of methodologies as discussed above. However, a cursory examination of the experimental literature in any developed part of science makes clear that there is a plethora of problems about the relation of evidence to theory that are not statistical in character and that need systematic analysis.

One task of formal analysis that I have tried to encourage several of my students to undertake, but as yet without any major success, is the detailed analysis of the relation between the major experimental evidence supporting quantum mechanics and the theory itself. There is, as far as I know, no place that systematically presents the evidence supporting the classical theory in a methodologically meticulous fashion. Indeed, it is the practice in textbooks and treatises on classical quantum mechanics not to present supporting experimental evidence in any serious form at all. A couple of years ago I wanted to determine how well the experimental evidence supported the claim that radioactive decay obeys an exponential probability law. I was surprised to find how difficult it was to locate in the literature a detailed statistical analysis of this problem and how much it was neglected even in its most superficial aspects in the standard discussions of radioactive decay. This, however, is a simple case. The much more complicated cases of the main experimental data supporting quantum mechanics are in a very unsatisfactory status from a systematic viewpoint, and much remains to be done.

To some extent, one of the best traditions in this respect is in psychology, but even here the formal theory of how experimental evidence is related to theory is not as explicit as it should be. I do want to emphasize the formidable character of the problems of organizing in a formal way the relationship between experimental evidence and theory. In recent years I have attempted to read some of the current experimental literature in physics. It is almost as if I had decided to learn a foreign language. Although I have a reasonable familiarity with certain areas of theoretical discussion in physics, I found the experimental literature to be an entirely different matter. The abbreviated technical descriptions of equipment, its functional characteristics, and the description of the data obtained, require a major effort of analysis to become conceptually independent of the large preceding literature on the same topic. I can now understand why my friends in physics tell me that the experimental literature in one area is almost unreadable even by someone working in a nearby area. I am not proposing that philosophers attempt to rework a large part of this literature—the task is clearly an impossible one. It would be of interest to have a detailed formal analysis of some particular areas of philosophical significance, for example, recent experiments on hidden-variable theories, if only for the purpose of bringing out how complicated the relation is between theory and experiment in the developed domains of science. The stories we get from formal accounts in philosophy unrelated to these details, or the equally simplified stories we get from historical accounts of past work in science when matters were technically much less complicated, are misleading. We need a corrective both to too much emphasis on purely formal methods on the one hand and purely historical methods on the other, by a proper and detailed look at current experiments in developed parts of science.

4. FINAL REMARK ON HISTORICAL AND FORMAL METHODS

There has been throughout this conference an obvious intellectual tension between those who advocate historical methods as the primary approach in the philosophy of science and those who advocate formal methods.

This tension in itself is a good thing. It generates both a proper spirit of criticism and a proper sense of perspective. Each group can tell the other about their weaknesses and the pursuit of philosophical matters can be undertaken at a deeper level. There is no worse fate for a developing theory or method than not to be confronted with opposing views that require a sharpening of concepts and a detailed development of arguments. On the other hand, there seems no reason not to find room in the philosophy of science for a vigorous pursuit of both historical and formal approaches.

I share with my more historically oriented colleagues a kind of horror at the thought of a formal philosophy of science that develops on its own, independent of the rich material offered by the sciences themselves. I have already mentioned some examples of this tendency and I join them in encouraging the pursuit of problems and of methods that have a complexity adequate to the actual work in developed sciences. But I also want to remind those concerned with historical methods in the philosophy of science that about ninety percent of all scientists who have ever lived are now alive, and the development of science since World War II is the most smashing success story in the history of thought. To be concerned only with the long historical perspective and not to understand the systematic details of modern science is as mistaken as the pursuit of empty formal methods that make no contact with developed scientific theories and their supporting experiments.

My intent is to end on the pluralistic note that there is more than enough interesting and important work for all of us. The tyranny of any single approach or any single method, whether formal or historical, should be vanquished by a democracy of methods that will coalesce and separate in a continually changing pattern as old problems fade away and new ones arise.

2

THE STUDY OF SCIENTIFIC REVOLUTIONS: THEORY AND METHODOLOGY

The nature of scientific revolutions has become a fashionable topic both in the history and in the philosophy of science. I shall not in this paper attempt to review the many controversies that have filled the literature in the past decade. My purpose is to try to take a longer view as to what the proper role of philosophy should be in the study of scientific revolutions. What I have to say is certainly tentative, and many of the ideas will probably be regarded as wrong by a fair number of my colleagues. This is not meant as a prefatory apology but simply as a prediction. I am quite prepared to defend what I have to say but regard the present study of such matters as so tentative and immature that to be at all certain of the correctness of my views would be too dogmatic for my skeptical and empirical view of philosophy, history, and science.

From the standpoint from which I approach the subject there are two natural divisions. One is consideration of the theory of scientific revolutions and the other is the methodology of evaluating the empirical soundness of such theories. In saying something about these matters I shall not make strenuous efforts to separate philosophy from other disciplines that can approach the same problems.

*Reprinted from, *La filosofía y las revoluciones científicas: Teoría y praxis.* México, D. F.: Editorial Grijalbo, S. A., 1979, pp. 295-306, but with the original English text.

1. THEORY

The theory of scientific revolutions seems to me to itself divide naturally into three parts. The first part is simply the description of the structure of science during the period that a presumed revolution took place. The second part concerns the kinematics of the presumed revolution, what is an appropriate description of the changes that took place, how can we describe those changes, and can we meaningfully talk about them as continuous or discrete in character? The third part concerns the dynamics of the revolution. Here the search is for causes and especially a *theory* of the causes.

Structure. As an example of the problems of having an adequate theory of structure, let us consider the history of geometry. Because of the paucity of texts, we can perhaps see in realistic terms the possibility of describing the state of geometry as a mathematical or scientific discipline in 200 B.C. in the Hellenistic world of Alexandria, Rhodes, Syracuse, and a few other places. Even then, the theory of what is to be regarded as essential in that structure and what is accidental or unimportant is not, as far as I can see, clearly formulated anywhere. Moreover, there are puzzles that seem difficult to solve in characterizing the structure; for example, how much relative weight should we attach to the methods of proof that were used as compared to the depth of the mathematical results that were obtained?

When we move across the centuries to the many rigorous formulations of geometry given at the end of the 19th century, with the work of Hilbert often being taken as a paradigm example, I at least find it even more perplexing to characterize what is to be regarded as the structure of the science of geometry. Are the rigorous axiomatic methods of Hilbert the most important feature, or is the group-theoretic viewpoint of Klein more fundamental? Moreover, this is to ask only the most elementary and primitive question. A structure as we ordinarily think of it is not properly characterized by simply listing its main features but rather by saying how these features are related and interlocked. We can talk with some precision about the structure of Euclidean Geometry in an abstract sense, but can we talk in a reasonably meaningful way about the structure of geometry as a scientific discipline at the end of the 19th century? What I have said about geometry seems to me to apply as well to any other major scientific discipline, running from astronomy to zoology.

The basis of the problem or, put another way, the reason for the absence of any substantive theory of structure is similar, it seems to me, to the absence of any systematic theory of structure for almost all historical phenomena of human interest. In the same way that we can question what is meant by the structure of a scientific discipline at a given time, we can ask about the structure of a society, the structure of a market, the structure of a military campaign. In those cases that have been regarded as of great historical interest it is fair to say that the concept of structure that is imposed is rudimentary at best and most often left at a completely impressionistic level. The reasons for this seem clear. We simply have not yet developed adequate abstractions to provide the basis for a serious theory of structure.

To my severe strictures about structure it is possible to reply that an unreasonable standard is being set, but it is important to recognize that here is a radical difference in that case between what we should hope to achieve in a given part of science itself as, for example, in the study of the structure of the atom, or the structure of the solar system, and what we have as our intellectual ambition about the structure of scientific revolutions. If the view is held that the theory of structure of such revolutions cannot rise above the present impressionistic state of affairs, then the theory of such matters will remain committed to a romantic view of what may be regarded as the highest products of our intellectual activities as human beings.

Kinematics. Without a theory of structure it is difficult to see how a theory of kinematics or of change can be developed. The kinematical theory of scientific revolution is in an even more primitive state than the theory of their structure. It is hard to think even of a nontrivial scientific problem that has yet been posed about such changes. Where indeed is to be found a testable theory or hypothesis of change about any branch of science?

Let us look at some typical questions that arise in the kinematics of a wide variety of natural phenomena and ask whether these questions can be transformed into meaningful programs of inquiry for the kinematics of scientific revolutions. One classical kind of question is whether change is continuous or discontinuous. The conservative postulate of most of classical physics, for example, is that change is continuous. In the case of classical mechanics, the further requirement would be imposed that the paths of particles are not merely continuous but piecewise twice differentiable. In the theory of Brownian motion, we end up with the result that the paths of particles are continuous but almost nowhere differentiable. On the other hand, in quantum mechanics a fundamental change in attitude was expressed in the discovery that the transitions between energy states of atoms were discontinuous and discrete rather than continuous. In the psychology of contemporary learning theory there has been an intense study of various kinds of learning, with some being characterized as continuous in character and others as being discrete or all-or-none. The differing theoretical assumptions that lead to these different kinematical predictions have been laid out in explicit detail.

These comparisons, it may be said, are unreasonable and unwarranted. Surely it is absurd, it may be claimed, to even think of distinguishing between differentiable and nondifferentiable continuous trajectories of change for scientific revolutions. With this point I agree, but the example from psychology illustrates the kind of theoretical work that can discriminate between our intuitive ideas of continuous versus discontinuous phenomena. In the case of the examples of learning, the main studies deal with experiments consisting of discrete trials, and there is no sharp notion of differentiability directly definable for the phenomena. On the other hand, very clear mutually contradictory hypotheses about the nature of learning can be formulated and given exact expression. Still, it may be said, the example is not appropriate because the study of historical phenomena cannot hope to achieve the precision or quantitative definiteness of developed experimental sciences.

But refuge in the nonexperimental character of historical phenomena is no refuge at all because for many centuries the most exact science, namely, astronomy, was and is wholly nonexperimental in character.

In rather brief and superficial terms it may be useful to make a comparison of an important but structurally simple kinematical theory of historical phenomena. I have in mind data on the modern rise in population. I take my discussion from McKeown (1976). Approximate estimates of the modern rise of population are given by McKeown as follows: By 1750 the world population is estimated to have been about 750 million; by 1830 it was one billion, two billion in 1930, three billion in 1960, and four billion in 1975. Through these data points we can fit a remarkably simple nonlinear function, and the kinematical theory consists of studying carefully which functions fit the data best. The example I have quoted is rather crude; much more exact population estimates exist for the more recent years and also for particular countries. In each case, the surface kinematical problem is to fit to the data a function that has a small number of parameters. For these kinds of data the problem is relatively simple. The kinematical problem faced by Kepler was not, as it was not for earlier Hellenistic astronomers, and especially if we regard, as we would today, the epicycle theory as a kinematical theory of the motion of the planets.

One problem about the kinematics of scientific revolutions that would be interesting to me, but perhaps would be regarded as not so by many philosophers and historians, would be the analysis of the rate of publication about a given topic across the period of a revolution. Do, for example, these spreads of publication have very similar mathematical form as a nearly universal characteristic of scientific revolutions, at least in the context of science since 1800, or are quite different rate functions to be found? I fear that my colleagues interested in the history of science and its relation to the philosophy of science have no real taste for such quantitative questions, but this, it seems to me, is mainly because they are not really interested in approaching their subject in a scientific fashion.

Dynamics. The population example cited earlier provides a good point to begin the discussion of dynamics. Many of us are alarmed at the nonlinear growth of population over the past hundred years, but of even greater interest is the investigation of the causes of this growth. As McKeown's book shows in some detail, a satisfactory causal analysis is not easy to come by, but some progress is possible and even some assessment of the contribution of modern medical discoveries and measures can be made. On the other hand, the theoretical status of causal analyses of major political upheavals such as the French or Russian revolutions seems to be in a shambles. An excellent survey was given some years ago by Howard K. Beale (1946) on the variety of attitudes toward causality and the nature of particular cause of the American Civil War. Here is how Beale summarizes some of the variety of views:

> Historians, whatever their predispositions, assign to the Civil War causes ranging from one simple force or phenomenon to patterns so complex and manifold that they include, intricately interwoven, all the important movements, thoughts, and actions of the decades before 1861. One writer finds in events of the immediately preceding years an adequate explanation of the War; another feels he must begin his story with 1831 or even 1820; still another goes back to the importation of the first slaves, to descriptions of geographic differences before white men appeared, or to differentiation in Europe between those who settled North and South.... Moral, ideological, political, economic, social, psychological explanations of the War have been offered. Responsibility has been ascribed both to action of men and to forces beyond human control. Conspiracy, constitutional interpretation, human wickedness, economic interest, divine will, political ambition, climate, "irrepressible conflict", emotion, rival cultures, high moral principles, and chance have severally been accredited with bringing on the War. There is a Marxian interpretation; also a racist theory. (pp. 55-56)

Beale goes on to spell out this vast variety of causal explanations, and any but the most dogmatic reader can scarcely end up with other than a highly skeptical attitude that it is possible at the present state of historical theory to provide a satisfactory or even partially satisfactory causal analysis of a major political or social revolution or conflict.

For much of the history of science, the development of a causal theory seems a futile exercise because of the paucity of data to test such a theory. This attitude is well expressed by Neugebauer (1957) in the following passage concerning causal theories of the origin of mathematics.

> The Greeks themselves had many theories about the origin of mathematics. A favored one, which is still kept alive in modern textbooks, makes the necessity of repeated land measurement responsible for geometry. Modern authors have often referred to the marvels of Egyptian architecture, though without ever mentioning a concrete problem of statics solvable by the known Egyptian arithmetical procedure. A much more sophisticated attitude is represented by Aristotle, who considers the existence of a "leisure class", to use a modern term, a necessary condition for scientific work. Our factual knowledge about the development of scientific thought and of the social position of the men who were responsible for it is so utterly fragmentary, however, that it seems to me completely impossible to test any such hypothesis, however plausible it may appear to a modern man. (pp. 151-152)

For science since 1800 or so, it may be felt that adequate data can be collected to test reasonable causal ideas, but, as the example of the American Civil War shows, we are faced in modern cases with the opposite difficulty namely, the data are so rich and varied that we have no serious idea as to how to make a scientific analysis of causes that can be properly defended.

Tales of detail from either an internalist or externalist standpoint about any particular scientific revolution are fascinating and intriguing to me as well as to many others, but I do not find in these lovely tales any trace of a serious scientific causal theory, and I am skeptical that in any near future we shall have one.

A proper role for philosophers here, as in other aspects of historical analysis, is to press the point about theory and to insist whether a commitment is being made or a claim is being made about the theoretical status of the propositions set forth on the nature of scientific revolutions. There is irony in the fact that after decades of formalist effort in the philosophy of science many philosophers seem to have been overcome by the richness of the data set in front of them by historians, no matter how primitive the theory that accompanies these data may be.

2. METHODOLOGY

In a number of sciences—experimental psychology and econometrics are perhaps the best examples—there is little development of the kind of detailed and rigorous theory I have been calling for. It may be thought by many that I am setting an unreasonable standard in drawing on developed theories in the physical sciences or in mathematics as models that should be followed in a theory of scientific revolutions. There is, of course, back of this issue a hoary problem of many years standing concerning the ideographic or nomothetic character of historical investigations. I am assuming without further debate that the case for the nomothetic view is overwhelming—at least it should be among those who want to make pretentious claims about the structure or the nature of scientific revolutions. My own attitude is plain: If the theory of scientific revolutions is primitive or nonexistent, let us not burke the facts.

But even if the theory is primitive, we can, as in the case of much of experimental psychology or econometrics, try to make serious scientific progress by application of a careful and explicitly thought out methodology. Some order can be brought to the welter of empirical data and some sense of cumulative progress can develop. A good many aspects of the historical study of population changes satisfy such a standard. Even that marvelous 18th century spinner of psychological fables, David Hume, was cautious, highly empirical, and careful in dealing with estimates of the population of the ancient world. The modern historical literature on population has become technical and scientific and to my mind all to its credit.

A valiant effort at developing a more quantitative methodology in the history of science has been made by Derek Price in his 1961 book and in a number of articles. Price has studied a number of phenomena of growth in science: the number of journals, the number of physics abstracts since 1900, the growth in the number of papers in a given field of science, and the growth in the number of scientists. He has investigated the extent to which exponential functions or other analytically simple nonlinear functions fit the data. He has not looked very much at scientific revolutions, but the kind of quantitative techniques he has begun to apply would not be inappropriate, especially in the analysis of the rise and fall of publications on a given scientific topic following its introduction into the literature. But in many ways Price has been a lonely example; not many people have followed the line of work he has begun. Above all, the detailed and tedious analysis of data required to pursue with any thoroughness the program he has started has not really taken place, and certainly not in the study of scientific revolutions. The result seems to be the inevitable one that the quantitative study of the history of science remains in a primitive state, just as the theory of scientific revolutions remains in such a state.

There is a notorious case of applying quantitative methods in history to which I would like to draw a parallel to what I think it would be desirable to see happen in the study of scientific revolutions or, more generally, in the study of many aspects of the history of science. In 1974, Robert William Fogel and Stanley L. Engerman published a controversial work, *Time on the Cross*, subtitled *The Economics of American Negro Slavery.* This work has become famous in the recent scholarship of American history for two reasons. First, it contravened a number of standard historical theses about the conditions of slavery and the performance of slaves in the pre-Civil War South. Second, the authors brought to bear as a method of establishing their theses a battery of statistical tools and techniques that have been developed and used extensively in econometrics but seldom, if at all, in the quantitative study of such matters as were the focus of their book. The repercussions of the work of Fogel and Engerman have been widespread in American circles of scholarship and there has perhaps been a tendency for a lineup of acceptance by historians oriented toward social science, on the one hand, and rejection by humanistically oriented historians on the other. But this is not the moral of my tale. A much more interesting outcome, in my judgment, is the painstaking and meticulous examination of the methodology of *Time on the Cross* by a group of economic historians sophisticated in the methods of econometrics. It is right and proper, in my view of things, that the really careful and exacting critique of Fogel and Engerman's work came from David, Gutman, Sutch, Temen, and Wright (1976), writing in the very spirit exemplified by *Time on the Cross* and not in terms of some humanistic broadside.

Sadly enough, the same kind of critical assessment and detailed analysis and reanalysis of data has not taken place within the framework begun by Price in the history of science. Compared with the sophistication of the methodology in *Time on the Cross* and the riposte of David et al. in *Reckoning with Slavery*, the quantitative methodology begun by Price is, especially from a statistical standpoint, still in its infancy. As Price remarks, "It is perhaps especially perverse of the historian of science to remain purely an historian and fail to bring the powers of science to bear upon the problems of its own structure. There should be much scope for scientific attack on science's own internal problems, yet, curiously enough, any such attack is regarded with much skepticism" (p. 93).

In the same year that *Time on the Cross* appeared, 1974, the distinguished American historian Eugene D. Genovese also published a book on American slavery entitled *Roll, Jordan, Roll: The World the Slaves Made.* Genovese writes in the traditional historical manner, giving his own intuitive digest of the vast amount of data surveyed, especially the personal accounts of the conditions of slavery in the old South. It is not my purpose here to assess the merits of Genovese's book, but one reviewer made a remark that I think is most appropriate. The really fundamental difference between *Time on the Cross* and *Roll, Jordan, Roll* is that the first can be proved wrong, and resoundingly so; the second is essentially inaccessible to either proof or disproof, for the methods do not lend themselves to any deeper analysis of evidence for or against any particular thesis. Fogel and Engerman made many mistakes but they were honest enough to lay out the data and to describe it in such a way that their tracks could be traced. Not so Genovese. It is not a question of intellectual dishonesty but a question of method. His tracks are covered not only from others but from himself. He cannot give a rational account of the methods by which his summary views or selections of individual sketches were made. I am happy to leave the creative sources of hypotheses or even of theories deep in the unconscious of the individual scientist or scholar but I am not happy at all to leave the methodology of verification at the same unconscious level. As far as I can see, this is where we still are in the analysis of scientific revolutions.

3

LIMITATIONS OF THE AXIOMATIC METHOD IN ANCIENT GREEK MATHEMATICAL SCIENCES

My thesis in this chapter is that the admiration many of us have for the rigor and relentlessness of the axiomatic method in Greek geometry has given us a misleading view of the role of this method in the broader framework of ancient Greek mathematical sciences. By stressing the limitations of the axiomatic method or, more explicitly, by stressing the limitations of the role played by the axiomatic method in Greek mathematical science, I do not mean in any way to denigrate what is conceptually one of the most important and far-reaching aspects of Greek mathematical thinking. I do want to emphasize the point that the use of mathematics in the mathematical sciences and in foundational sciences, like astronomy, compare rather closely with the contemporary situation. It has been remarked by many people that modern physics is by and large scarcely a rigorous mathematical subject and, above all, certainly not one that proceeds primarily by extensive use of formal axiomatic methods. It is also often commented upon that the mathematical rigor of contemporary mathematical physics,

*Reprinted from *Theory change, ancient axiomatics, and Galileo's methodology, Pisa Conference Proceedings* Vol. 1 (ed. by J. Hintikka, D. Gruender and E. Agazzi), 1980, pp. 197–213.

in relation to the standards of rigor in pure mathematics today, is much lower than was characteristic of the 19th century. However, my point about the axiomatic method applies also to 19th-century physics. There is little evidence of rigorous use of axiomatic methods in that century either. This is true not only of the periodical literature but also of the great treatises. Three casual examples that come to mind are Laplace's *Celestial Mechanics*, his treatise on probability, and Maxwell's treatise on electricity and magnetism.

Three examples from ancient Greek mathematical sciences that I have chosen to comment on are Euclid's *Optics*, Archimedes' *On the Equilibrium of Planes*, and Ptolemy's *Almagest.*

1. EUCLID'S OPTICS

It is important to emphasize that Euclid's *Optics* is really a theory of vision and not a treatise on physical optics. A large number of the propositions are concerned with vision from the standpoint of perspective in monocular vision. Indeed, Euclid's *Optics* could be characterized as a treatise on perspective within Euclidean geometry. The tone of Euclid's treatise can be seen from quoting the initial part, which consists of seven 'definitions'.

1. Let it be assumed that lines drawn directly from the eye pass through a space of great extent;

2. and that the form of the space included within our vision is a cone, with its apex in the eye and its base at the limits of our vision;

3. and that those things upon which the vision falls are seen, and that those things upon which the vision does not fall are not seen;

4. and that those things seen within a larger angle appear larger, and those seen within a smaller angle appear smaller, and those seen within equal angles appear to be of the same size;

5. and that those things seen within the higher visual range appear higher, while those within the lower range appear lower;

6. and, similarly, that those seen within the visual range on the right appear on the right, while those within that on the left appear on the left;

7. but that things seen within several angles appear to be more clear.

(The translation is taken from that given by Burton in 1945.)

The development of Euclid's *Optics* is mathematical in character, but it is not axiomatic in the same way that the *Elements* are. For example, Euclid later proves two propositions, "to know how great is a given elevation when the sun is shining" and "to know how great is a given elevation when the sun is not shining". As would be expected, there is no serious introduction of the concept of the sun or of shining but they are treated in an informal, commonsense, physical way with the essential thing for the proof being rays from the sun falling upon the end of a line. Visual space is of course treated by Euclid as Euclidean in character.

It might be objected that there are similar formal failings in Euclid's *Elements*, but it does not take much reflection to recognize the very great difference between the introduction of many sorts of physical terms in these definitions from the *Optics* and the very restrained use of language to be found in the *Elements*. Moreover, the proofs have a similar highly informal character. It seems to me that the formulation of fundamental assumptions in Euclid's *Optics* is very much in the spirit of what has come to be called, in our own time, physical axiomatics. There is no attempt at any sort of mathematical rigor but an effort to convey intuitively the underlying assumptions.[1]

2. ARCHIMEDES' ON THE EQUILIBRIUM OF PLANES

Because I want to discuss the Archimedean treatise in some detail, a review of the theory of conjoint measurement is needed. The mixture of highly explicit axioms of conjoint measurement (as we would call them) and very inexplicit axioms about centers of gravity make Archimedes' treatise a peculiarly interesting example.

Conjoint measurement. In many kinds of experimental or observational environments, the measurement of a single magnitude of property is not feasible or theoretically interesting. What is of interest, however, is the joint measurement of several properties simultaneously. The intended representation is that we consider ordered pairs of objects or stimuli. The first members of the pairs are drawn from one set, say A_1, and consequently represent one kind of property or magnitude; the second members

[1] Ptolemy's *Optics* is much more physical and experimental in character. A more mathematical example, without any explicit axioms at all, is Diocles' treatise *On Burning Mirrors* (Toomer, 1976). The detailed mathematical proofs are also interesting in Diocles' work because of the absence in most cases of reasons justifying the steps in the argument, but, as in a modern nonaxiomatic text, familiar mathematical facts and theorems are used without comment.

of the pairs are objects drawn from a second set, say A_2, and represent a different magnitude or property. Given the ordered pair structure, we shall only require judgments of whether or not one pair jointly has more of the 'conjoined' attribute than a second pair.

Examples of interpretations for this way of looking at ordered pairs are abundant. In Archimedes' case, we are dealing with the measurement of static moments of force, or torques, where the two properties that make up the conjoint attribute are mass (or weight) and distance from the fulcrum. Momentum is another familiar example of a conjoint attribute. Quite different examples may be drawn from psychology or economics. For instance, a pair (a,p) can represent a tone with intensity a and frequency p, and the problem is to judge which of the two tones sounds louder. Thus the individual judges $(a,p) \succeq (b,q)$ if and only if tone (a,p) seems at least as loud as (b,q).

The axioms of conjoint measurement are stated in terms of a single binary relation defined on the Cartesian product $A_1 \times A_2$. All the axioms have an elementary character, except for the Archimedean axiom, which I shall not formulate explicitly along with the other axioms, but which I discuss below. In formulating the axioms, I use the usual equivalence relation $\approx$, which is defined in terms of $\succeq$, i.e., $(a,p) \approx (b,q)$ if and only if $(a,p) \succeq (b,q)$ and $(b,q) \succeq (a,p)$. Later, we shall also use the strict ordering: $(a,p) \succ (b,q)$ if and only if $(a,p) \succeq (b,q)$ and not $(b,q) \succeq (a,p)$. The axioms are embodied in the following definition.

DEFINITION 1. *A structure* $\langle A_1, A_2, \succeq \rangle$ *is a* conjoint structure *if and only if the following axioms are satisfied for every a, b and c in* A_1 *and every p, q and r in* A_2:

Axiom 1. If $(a,p) \succeq (b,q)$ *and* $(b,q) \succeq (c,r)$ *then* $(a,p) \succeq (c,r)$;

Axiom 2. $(a,p) \succeq (b,q)$ *or* $(b,q) \succeq (a,p)$;

Axiom 3. If $(a,p) \succeq (b,p)$ *then* $(a,q) \succeq (b,q)$;

Axiom 4. If $(a,p) \succeq (a,q)$ *then* $(b,p) \succeq (b,q)$;

Axiom 5. If $(a,p) \succeq (b,q)$ *and* $(b,r) \succeq (c,p)$ *then* $(a,r) \succeq (c,q)$;

Axiom 6. There is an s in A_2 *such that* $(a,p) \approx (b,s)$;

Axiom 7. There is a d in A_1 such that $(a,p) \approx (d,q)$;

Axiom 8. Archimedean axiom.

The intuitive content of most of the axioms is apparent. Axiom 1 is merely the familiar requirement of transitivity and Axiom 2 that of strong connectivity. Axioms 3 and 4 express the independence of one component from the other. Axioms 3 and 4 actually follow from the other axioms, but in the treatment of Krantz, Luce, Suppes, and Tversky (1971), weaker solvability axioms are used than Axioms 6 and 7, and in that context, Axioms 3 and 4 are needed. In any case, they state an important conceptual property. Axiom 5 states a cancellation property. When it is formulated in terms of the equivalence relation $\approx$ instead of $\succeq$, it is called the Thomsen condition, especially in the theory of webs. As already remarked, Axioms 6 and 7 state simple solvability axioms. Finally, Axiom 8 must be some form of the Archimedean axiom. Of course, I mean not an axiom directly pertinent to the treatise we are discussing here, but the familiar Archimedean axiom which is usually attributed to Eudoxus and not to Archimedes. In its most familiar form, it says that if we are given two magnitudes and the first is less than the second, there is a finite multiple of the first that is larger than the second. To formulate the axiom in explicit mathematical form in the present context, with no concept of addition or multiplication directly given, is somewhat troublesome. Because it is not important for our present discussion, I shall leave the axiom in inexplicit form.

For subsequent discussion of the postulates stated in Archimedes' treatise, some elementary consequences of Axioms 1–4 of Definition 1 are useful.

THEOREM 1. *The relation $\approx$ is an equivalence relation on $A_1 \times A_2$, i.e., it is reflexive, symmetric and transitive on $A_1 \times A_2$; and the relation $\succ$ is irreflexive, asymmetric and transitive on $A_1 \times A_2$.*

It is also desirable to define corresponding relations for each component.

Thus, for a and b in $A_1, a \succeq_1 b$ if and only if for some p in $A_2, (a,p) \succeq (b,p)$; and for p and q in $A_2, p \succeq_2 q$ if and only if for some a in $A_1, (a,p) \succeq (a,q)$. Then as before, we may define for $i = 1,2, x \approx_i y$ if and only if $x \succeq_i y$ and $y \succeq_i x$; and $x \succ_i y$ if and only if $x \succeq_i y$ and not $y \succeq_i x$. Using especially Axioms 3 and 4, the independence axioms, we may easily prove the following theorem.

THEOREM 2. *For $i = 1, 2$, the relation $\succeq_i$ is transitive and strongly connected on A_i, the relation $\approx_i$ is an equivalence relation on A_i and the* *relation $\succ_i$ is irreflexive, asymmetric and transitive on A_i.*

We can prove that any structure satisfying the axioms of Definition 1 can be given either an additive or a multiplicative representation in terms of real numbers. Because the multiplicative representation is most pertinent here, we shall state the basic representation theorem in that form. The reader is referred to Krantz et al. (1971, Chapter 6) for the proof of the theorem.

THEOREM 3. *Let $\langle A_1, A_2, \succeq\rangle$ be a conjoint structure. Then there exist real-valued functions φ_1 and φ_2 on A_1 and A_2, respectively, such that for a and b in A_1 and p and q in A_2*

$$\varphi_1(a)\varphi_2(p) \geq \varphi_1(b)\varphi_2(q) \textit{ if and only if } (a,p) \succeq (b,q).$$

Moreover, if φ_1' and φ_2' are any two other functions with the same property, then there exist real numbers $\alpha, \beta_2, \beta_2 > 0$ such that

$$\varphi_1 = \beta_1\varphi_1^\alpha$$

and

$$\varphi_2 = \beta_2\varphi_2^\alpha,$$

provided there are elements a and b in A_1 and p in A_2 such that $(a,p) \succ (b,p)$, and elements p and q in A_2 and c in A_1 such that $(c,p) \succ (c,q)$.

More than the theory of conjoint measurement is needed to give a correct analysis of Archimedes' treatise, for he obviously assumes that weight and distance are extensive or additive magnitudes. (This point is documented in the later discussion.) It will therefore also be useful to have in front of us the modern theory of extensive magnitudes. A rather complete presentation of the theory is to be found in Krantz et al. (1971, Chapter 3). Because of their relative simplicity I shall state here the axioms of Suppes (1951). In this case the Archimedean axiom is easily stated explicitly. A binary operation $\circ$ on the set A of magnitudes, as well as a binary relation $\succeq$, is introduced, and we define recursively $1x = x$ and $nx = (n-1)x \circ x$. As before, the relations $\approx$ and $\succ$ are defined as expected in terms of $\succeq$.

DEFINITION 2. *A structure $\langle A, \succeq, \circ\rangle$ is a* structure of extensive magnitudes *if and only if the following axioms are satisfied for every a, b and c in A:*

Axiom 1. If $x \succeq y$ and $y \succeq z$ then $x \succeq z$;

Axiom 2. $(x \circ y) \circ z \succeq x \circ (y \circ z)$;

Axiom 3. If $x \succeq y$ *then* $x \circ z \succeq z \circ y$;

Axiom 4. If $x \succ y$ *then there is a* z *in* A *such that* $x \approx y \circ z$;

Axiom 5. $x \circ y \succ x$;

Axiom 6. If $x \succeq y$ *then there is a natural number* n *such that* $y \succeq nx$.

The six axioms of Definition 2 have an obvious content when A is a set of positive numbers closed under addition and subtraction of smaller numbers from larger ones, $\succeq$ is the numerical weak inequality, and $\circ$ is the operation of addition. It should be noted that Axiom 3 combines monotonicity and commutativity. The numerical interpretation just given is itself the basis of the following representation theorem.

THEOREM 4. *Let* $\langle A, \succeq, \circ \rangle$ *be a structure of extensive magnitudes. Then there exists a real-valued function* φ *on* A *such that for* a *and* b *in* A

$$\varphi(a) \geq \varphi(b) \text{ if and only if } a \succeq b,$$

and

$$\varphi(a \circ b) = \varphi(a) + \varphi(b).$$

Moreover, if φ' *is any other such function then there is a real number* $a > 0$ *such that* $\varphi' = \alpha\varphi$.

Archimedes' postulates. With the axioms of conjoint and extensive measurement given above as background, let us now turn to Archimedes' postulates at the beginning of Book I of *On the Equilibrium of Planes.* I cite the Heath translation.

> I postulate the following:
>
> 1. Equal weights at equal distances are in equilibrium, and equal weights at unequal distances are not in equilibrium but incline towards the weight which is at the greater distance.
>
> 2. If, when weights at certain distances are in equilibrium, something be added to one of the weights, they are not in equilibrium but incline towards that weight to which the addition was made.
>
> 3. Similarly, if anything be taken away from one of the weights, they are not in equilibrium but incline towards the weight from which nothing was taken.

> 4. When equal and similar plane figures coincide if applied to one another, their centers of gravity similarly coincide.
>
> 5. In figures which are unequal but similar the centers of gravity will be similarly situated. By points similarly situated in relation to similar figures I mean points such that, if straight lines be drawn from them to the equal angles, they made equal angles with the corresponding sides.
>
> 6. If magnitudes at certain distances be in equilibrium, (other) magnitudes equal to them will also be in equilibrium at the same distances.
>
> 7. In any figure whose perimeter is concave in (one and) the same direction the centre of gravity must be within the figure.

Looking at the postulates, it is clear that postulates 1, 2, 3 and 6 fall within the general conceptual framework of conjoint measurement, but the remaining postulates introduce geometrical ideas that go beyond the general theory of conjoint measurement. I shall have something more to say about these geometrical postulates later. For the moment I want to concentrate on what I have termed the *conjoint postulates.* The wording of Postulates 2 and 3 makes it clear that Archimedes treated weight as an extensive magnitude. We shall thus assume that $\mathcal{W} = \langle W, \succeq_1, \circ \rangle$ is a structure of extensive magnitudes, that $\langle W \times D, \succeq \rangle$ is a conjoint structure, and that $\succeq_1$ of $\mathcal{W}$ is the defined relation $\succeq_1$, of the conjoint structure. Also, to formulate Postulate 3 explicitly we need a subtraction operation that is well defined for extensive structures: If $x \succ y$ then $x - y \approx z$ if and only if $x \approx y \circ z$.

The formulation of Postulates 1, 2,3 and 6 then assumes the following elementary form, with subscripts of $\succeq_1$ and $\succeq_2$ dropped to simplify the notation.

1a. *If* $w_1 \approx w_2$ *and* $d_1 \approx d_2$ *then* $(w_1, d_1) \approx (w_2, d_2)$.

1b. *If* $w_1 \approx w_2$ *and* $d_1 \succ d_2$ *then* $(w_1, d_1) \succ (w_2, d_2)$.

2. *If* $(w_1, d_1) \approx (w_2, d_2)$ *then* $(w_1 \circ x, d_1) \succ (w_2, d_2)$.

3. *If* $(w_1, d_1) \approx (w_2, d_2)$ *and* $w_2 \succ x$ *then* $(w_1, d_1) \succ (w_2 - x, d_2)$.

6. *If* $(w_1, d_1) \approx (w_2, d_2), w_3 \approx w_1$ *and* $w_4 \approx w_2$ *then* $(w_3, d_1) \approx (w_4, d_2)$.

The first three propositions of Book I can be proved from these purely conjoint postulates and the assumption that weight is an extensive magnitude. For detailed analysis I cite the Heath translation of the propositions and their proofs.

Proposition 1.

> *Weights which balance at equal distances are equal.*
> For, if they are unequal, take away from the greater the difference between the two. The remainders will then not balance [*Post.* 3]; which is absurd.
> Therefore the weights cannot be unequal.

Proposition 2.

> *Unequal weights at equal distances will not balance but will incline towards the greater weight.*
> For take away from the greater the difference between the two. The equal remainders will therefore balance [*Post.* 1]. Hence, if we add the difference again, the weights will not balance but incline towards the greater [*Post.* 2].

Proposition 3.

> *Unequal weights will balance at unequal distances, the greater weight being at the lesser distance.*
> Let A, B be two unequal weights (of which A is the greater) balancing about C at distances AC, BC respectively.
> Then shall AC be less than BC. For, if not, take away from A the weight $(A - B)$. The remainders will then incline towards B [Post. 3]. But this is impossible, for (1) if $AC = CB$, the equal remainders will balance, or (2) if $AC > CB$, they will incline towards A at the greater distance [Post. 1].
> Hence $AC < CB$.
> *Conversely*, if the weights balance, and $AC < CB$, then $A > B$.

My aim is to catch the spirit of Archimedes' formulation of these first three propositions *and* their proofs within the formalization I have given. To be as explicit as possible about my procedure, I use in the proofs elementary properties of extensive magnitudes that follow from the axioms of Definition 2, but only properties of conjoint structures that follow from Archimedes' postulates, not the full set of Definition 1.

PROPOSITION 1. *If* $(w_1, d_1) \approx (w_2, d_2)$ *and* $d_1 \approx d_2$ *then* $w_1 \approx w_2$.

Proof. Suppose $w_1 \succ w_2$. Let $z = w_1 - w_2$. Then $w_1 - z \approx w_2$. Then by Postulate 1a

$$(w_1 - z, d_1) \approx (w_2, d_2), \tag{1}$$

but by Postulate 3 and the hypothesis of the theorem

(2) $$(w_2, d_2) \succ (w_1 - z, d_1),$$

and (1) and (2) are from the definitions of $\succ$ and $\approx$ jointly absurd.

PROPOSITION 2. *If* $w_1 \succ w_2$ *and* $d_1 \approx d_2$ *then* $(w_1, d_1) \succ (w_2, d_2)$.

Proof. Let $z = w_1 - w_2$. Then $w_1 - z \approx w_2$, and by Postulate 1a

$$(w_1 - z, d_1) \approx (w_2, d_2).$$

Therefore, by Postulate 2

$$((w_1 - z_1) \circ z_1, d_1) \succ (w_2, d_2),$$

and $(w_1 - z_1) \circ z_1 = w_1$, so

$$(w_1, d_1) \succ (w_2, d_2).$$

PROPOSITION 3. *If* $w_1 \succ w_2$ *and* $(w_1, d_1) \approx (w_2, d_2)$ *then* $d_2 \succ d_1$.

Proof. Suppose not $d_2 \succ d_1$. Let $z = w_1 - w_2$. Then by Postulate 3,

(1) $$(w_2, d_2) \succ (w_1 - z, d_1),$$

but this we shall show is absurd. First if $d_1 \approx d_2$, then by Postulate 1a

(2) $$(w_1 - z_1, d_1) \approx (w_2, d_2),$$

and (as in the proof of Prop. 1) (1) and (2) are jointly absurd. On the other hand, if $d_1 \succ d_2$, then by Postulate 1b

(3) $$(w_1 - z_1, d_1) \succ (w_2, d_2),$$

and (1) and (3) are jointly absurd (from the asymmetry of $\succ$). Hence $d_2 \succ d_1$.

On one point my formalization is clearly not faithful to Archimedes. I have replaced his symmetrical relation *unequal* by the asymmetric $\succ$, but this is a trivial formal difference, easy to eliminate if desired.

The remaining propositions of Book I use the concept of center of gravity in either their formulations or proofs, and I defer the consideration of this much-disputed concept.

The postulates and propositions as I have reformulated them above are a part of the elementary theory of conjoint measurement on the assumption that the first component is a structure of extensive magnitudes as well. A casual perusal of modern textbooks on mechanics reveals quickly enough that postulates like the ones formulated here are not an explicit part of modern discussions of static moments of force. The reason is simple. Once a numerical representation is assumed, explicit conjoint axioms are not necessary. Take Postulate 1a, for instance, and use the multiplicative representation:

$$\text{If } \varphi_1(w_1) = \varphi_1(w_2) \text{ and } \varphi_2(d_1) = \varphi(d_2) \text{ then}$$
$$\varphi_1(w_1)\varphi_2(d_1) = \varphi_1(w_2)\varphi_2(d_2),$$

but this is just an elementary truth of arithmetic and consequently not necessary to assume.

The important historical fact is that the concept of a numerical representation was missing in Greek mathematics, and consequently explicit conjoint axioms were needed. There seems little doubt that Archimedes' statement of such axioms is historically the earliest instance of an explicit approach to conjoint measurement, certainly at least in terms of extant texts of Greek mathematics and science.

It has been noted by many modern commentators that Greek mathematicians were completely at ease in comparing ratios of different sorts of magnitudes, e.g., the ratio of two line segments to that of two areas. Given this tradition it is natural to query why Archimedes did not state the Postulates of Book I in terms of ratios. The answer it seems to me is clear. Proof that two weights balance at distances reciprocally proportional to their magnitudes, which is Propositions 6 and 7 of Book I, is the Greek equivalent of a numerical representation theorem in the theory of measurement. The conjoint postulates that Archimedes formulates provide a simple qualitative basis from which the Greek 'representation theorem' can be proved. (I shall have more to say later about this proof.)

I know of no other instance of conjoint concepts in Greek mathematics and science. Certainly modern examples like momentum were not considered, and no such concepts were needed in Archimedes' other physical work, *On Floating Bodies.* It is perhaps for this reason that the level of abstraction to be found, for example, in Book V of Euclid's *Elements* is not reached in *On the Equilibrium of Planes.*[2] A higher level of abstraction was superfluous because other pairs of magnitudes satisfying like postulates were not known.

[2] The attitude toward abstraction is very clearly expressed by Aristotle in the *Posterior Analytics* (Book I, 5, 74a 17-25). "An instance of (2) would be the law that proportionals alternate. Alternation used to be demonstrated separately of numbers, lines, solids, and durations, though it could have been proved of them all by a single demonstration. Because there was no single name to denote that in which numbers, lengths, durations, and solids are identical, and because they differed specifically from one another, this property was proved of each of them separately. Today, however, the proof is commensurately universal, for they do not possess this attribute *qua* lines or *qua* numbers, but *qua* manifesting this generic character which they are postulated as possessing universally". The reference to (2) is to one kind of error we can make in drawing a conclusion that is too specific or concrete. Errors of type (2) arise "when the subjects belong to different species and there is a higher universal, but it has no name" (74a 7).

Centers of gravity. The most difficult conceptual problem of Archimedes' treatises concerns the status of the concept of center of gravity of a plane figure. This concept is essential to the formulation of Postulates 4, 5 and 7, but it is quite evident, on the other hand, that these postulates in themselves do not provide a complete characterization of the concept. By this I mean that if we knew nothing about centers of gravity except what is stated in Postulates 4, 5 and 7, we would not be able to derive the theorems in which Archimedes is interested, and which he does derive. As Dijksterhuis (1956) points out, it is possible to argue that the concept of center of gravity is being taken over by Archimedes from more elementary discussions and thus really has the same status as the geometrical concept of similarity in his treatise. On the face of it, this argument seems sounder than that of Toeplitz and Stein (published in Stein, 1930), who propose that the postulates are to be taken as implicitly defining centers of gravity once the postulates are enlarged by the obvious and natural assumptions.

It is also clear that a standard formalization of Archimedes' theory, in the sense of first-order logic, cannot be given in any simple or elegant way. It is possible to give the standard formalization of the part of the theory embodied in Postulates 1, 2, 3 and 6, as we have seen in the previous section.

Quite apart from the question of standard formalization, there are serious problems involved in giving a reconstruction in set-theoretical terms of Archimedes' postulates. In such a set-theoretical formulation, we can without difficulty use a geometrical notion like similarity. If we take over from prior developments a definition of center of gravity, then it would seem that Postulate 4, for example, would simply be a theorem from these earlier developments and would not need separate statement. Put another way, under this treatment of the concept of center of gravity, no primitive notion of Archimedes' theory would appear in Postulate 4 and thus it would clearly be an eliminable postulate. The same remarks apply to Postulates 5 and 7. It would seem that Archimedes has constructed a sort of halfway house; his postulates do not give a complete characterization of centers of gravity, but on the other hand, they cannot be said to depend upon a completely independent characterization of this concept.

Schmidt (1975) gives an interesting axiomatic reconstruction of Archimedes' theory, but his elegant postulates for centers of gravity are restricted to plane polygonal figures, whereas in Book II Archimedes is especially concerned with centers of gravity of parabolic segments. The 'reduction' of such segments to rectangles of equal area requires the results found in Archimedes' treatise *Quadrature of the Parabola.* (Schmidt's treatment of the 'conjoint' axioms discussed above does not use the standard modern results on conjoint measurement.)

It is worth noting that the fundamental pair of propositions (6 and 7) asserting the law of the lever, or what we may also term the law of static torque, does not really need any geometrical facts about centers of gravity, as do later propositions of Book I, and the whole of Book II. Archimedes could have used something like the following definition to get as far as Proposition 7: *The center of gravity of* (w_1, d_1) *and* (w_2, d_2) *is the distance* d_3 *such that* $(w_1, d_2 - d_3) \approx (w_2, d_3 - d_1)$. This definition assumes that distances are extensive magnitudes, but there is little difficulty about this assumption. It seems obvious to me why it is unlikely Archimedes even momentarily would have considered such a definition. The mathematically difficult and geometrically significant propositions all deal with the centers of gravity of geometric figures; in fact, the whole of Book II is concerned with finding the centers of gravity of parabolic segments, and for this purpose a geometric concept of center of gravity is a necessity.[3]

From a purely axiomatic standpoint, therefore, Archimedes is no more satisfactory than a modern physical treatise with some mathematical pretensions. A good comparative example, perhaps, is von Neumann's book (1932/1955) on quantum mechanics, which contains a beautifully clear axiomatic development of the theory of Hilbert spaces, but not of quantum mechanics itself.

[3] In closing this discussion, it is worth noting that Mach (1942), in his famous treatise on mechanics, seems to be badly confused on what Archimedes' work is all about. The focus of Mach's analysis is the famous Proposition 6 asserting that commensurable magnitudes are in equilibrium at distances reciprocally proportional to their weights. Mach is particularly exercised by the fact that "the entire deduction (of this proposition) contains the proposition to be demonstrated by assumption if not explicitly" (p. 20). A central point of Mach's confusion seems to be a complete misunderstanding as to the nature of the application of mathematics to physics. He seems to have no real conception of how mathematics is used to derive particular propositions from general assumptions, and what the relation of these general assumptions to the particular proposition is. He seems to think that any such proposition as the one just quoted must somehow be established directly from experience. His mistaken sentiments on these matters are clearly expressed in the following passage:

> From the mere assumption of the equilibrium of equal weights at equal distances is derived the inverse proportionality of weight and lever arm! How is that possible? If we were unable philosophically and a priori to excogitate the simple fact of the dependence of equilibrium on weight and distance, but were obliged to go for that result to experience, in how much less a degree shall we be able, by speculative methods, to discover the form of this dependence, the proportionality! (p. 19)

This last quotation shows, it seems to me, the basic fact that is usually not explicitly admitted in discussing Mach's views on the foundations of mechanics. He simply had no coherent or reasonable conception of how mathematics can be used in science, and his wrong-headed analysis of Archimedes is but one of many instances that support this conclusion.

3. PTOLEMY'S ALMAGEST

The third and most important example I cite is Ptolemy's *Almagest.* It is significant because it is the most important scientific treatise of ancient times and because it does not contain any pretense of an axiomatic treatment.

It is to be emphasized that Ptolemy uses mathematical argument, and indeed mathematical proof, with great facility, but he uses the mathematics in an applied way. He does not introduce explicit axioms about the motion of stellar bodies, but reduces the study of their motion to geometrical propositions, including of course the important case of spherical trigonometry.

Near the beginning of the *Almagest*, Ptolemy illustrates very well in the following passage the spirit of the way in which assumptions are brought in:[4]

> And so in general we have to state that the heavens are spherical and move spherically, that the earth in figure is also spherical to the senses when taken in all its parts; in position lies right in the middle of the heavens, like a geometrical center; and in magnitude and distance has the ratio of a point with respect to the sphere of the fixed stars, having no local motion itself at all. And we shall go through each of these points briefly to bring them to mind (p. 7).

There then follows a longer and more detailed discussion of each of these matters, such as the proposition that the heavens move spherically. My point is that the discussion and the framework of discussion are very much in the spirit of what we think of as nonaxiomatic mathematical sciences today. There is not a hint of organizing these ideas in axiomatic fashion. When Ptolemy gets down to details he has the following to say:

> But now we are going to begin the detailed proofs. And we think the first of these is that by means of which is calculated the length of the arc between the poles of the equator and the ecliptic, and which lies on the circle drawn through these poles. To this end we must first see expounded the method of computing the values of chords inscribed in a circle, which we are now going to prove geometrically, once for all, one by one (p. 14).

[4] The quotations given here are adapted from the translation by Taliaferro (1952), but after this article was written the definitive English translation by Toomer (1984) appeared, which will be standard reference in English for many years. The two passages cited do not differ materially from Toomer's.

The detailed discussion, then, on the size of chords inscribed in a circle emphasizes, above all, calculation and would make a modern physicist happy by its tone and results as well. This long and important analysis of computations is concluded with a numerical table of chords.

The thesis I am advancing is illustrated, in many ways even more strikingly, by the treatment of the motion of the moon in Book IV. Here Ptolemy is concerned to discuss in considerable detail the kind of observations that are appropriate for a study of the moon's motion and especially with the methodology of how a variety of observations are to be rectified and put into a single coherent theory.

Various hypotheses introduced in later books, e.g., the hypothesis of the moon's double anomaly in Book V, are in the spirit of modern astronomy or physics, not axiomatic mathematics. Moreover, throughout the *Almagest*, Ptolemy's free and effective use of geometrical theorems and proofs seems extraordinarily similar in spirit to the use of the differential and integral calculus and the theory of differential equations in a modern treatise on some area of mathematical physics.

4. CONCLUDING REMARKS

In this analysis of the use of axiomatic methods and their absence in explicit form in ancient mathematical sciences such as optics and astronomy, I have not entered into a discussion of the philosophical analysis of the status of axioms, postulates and hypotheses. There is a substantial ancient literature on these matters running from Plato to Proclus. Perhaps the best and most serious extant discussion is to be found in Aristotle's *Posterior Analytics*. Aristotle explains in a very clear and persuasive way how geometrical proofs can be appropriately used in mechanics or optics (75b 14ff). But just as Aristotle does not really have any developed examples from optics, mechanics or astronomy, so it seems to me that the interesting distinctions he makes do not help us understand any better the viewpoint of Euclid toward the 'definitions' of his optics or the postulates of Archimedes about centers of gravity cited above.

Many of you know a great deal more than I do about the history of Greek mathematics and Greek mathematical sciences, but, all the same, I want to venture my own view of the situation I have been describing. I may be too much influenced by my views about contemporary science, but I find little difference between contemporary physics and the problems of Greek science I have been describing. Physicists of today no more conform to an exact canon of philosophical analysis in their setting forth of physical principles or ideas than did those ancient scientists and mathematicians who wrote about the subjects I have been discussing. There was certainly a sense of methodology deeply embedded in Euclid, Archimedes and Ptolemy, but it was not a sense of methodology that was completely explicit or totally worked out, just as Aristotle's own general principles are never exemplified in any detailed and complicated scientific examples of an extended sort. The gap between philosophical analysis, canons of axiomatic method, and actual working practice was about the same order of magnitude that it is today. What is surprising, I think, from a philosophical standpoint is that the gap seems, if anything, to have widened rather than narrowed over the past 2000 years.

4

THE PLURALITY OF SCIENCE

What I have to say falls under four headings: What is unity of science, unity and reductionism, the search for certainty, and the search for completeness.

1. WHAT IS UNITY OF SCIENCE SUPPOSED TO BE?

To answer this initial question, I turned to the introductory essay by Otto Neurath (1938) for Volume 1, Part 1, of the *International Encyclopedia of Unified Science.* He begins this way:

> Unified science became historically the subject of this Encyclopedia as a result of the efforts of the unity of science movement, which includes scientists and persons interested in science who are conscious of the importance of a universal scientific attitude.
>
> The new version of the idea of unified science is created by the confluence of divergent intellectual currents. Empirical work of scientists was often antagonistic to the logical constructions of a priori rationalism bred by philosophico-religious systems; therefore,"empiricalization" and "logicalization" were consid-

*Reprinted from *PSA 1978*, Vol 2 (ed. by P. Asquith and I. Hacking), 1981, 3-16. East Lansing MI: Philosophy of Science Association. I am indebted to Georg Kreisel for a number of penetrating criticisms of the first draft of this paper.

> ered mostly to be in opposition—the two have now become synthesized for the first time in history (1938, p. 1).

Later he continues:

> All-embracing vision and thought is an old desire of humanity.... This interest in combining concepts and statements without empirical testing prepared a certain attitude which appeared in the following ages as metaphysical construction. The neglect of testing facts and using observation statements in connection with all systematized ideas is especially found in the different idealistic systems (1938, pp. 5-6).

Later he says:

> A universal application of logical analysis and construction to science in general was prepared not only by empirical procedure and the systematization of logico-empirical analysis of scientific statements, but also by the analysis of language from different points of view (1938, pp. 16-17).

In the same volume of the *Encyclopedia*, the thesis about the unity of the language of science is taken up in considerably more detail in Carnap's analysis of the logical foundations of the unity of science. He states his well-known views about physicalism and, concerning the terms or predicates of the language, concludes:

> The result of our analysis is that the class of observable thing-predicates is a sufficient reduction basis for the whole of the language of science, including the cognitive part of the everyday language (1938, p. 60).

Concerning the unity of laws, Carnap reaches a negative but optimistic conclusion—optimistic in the sense that the reducibility of the laws of one science to another has not been shown to be impossible. Here is what he has to say on the reduction of biological to physical laws:

> There is a common language to which both the biological and the physical laws belong so that they can be logically compared and connected. We can ask whether or not a certain biological law is compatible with the system of physical laws, and whether or not it is derivable from them. But the answer to these questions cannot be inferred from the reducibility of

> the terms. At the present state of the development of science, it is certainly not possible to derive the biological laws from the physical ones. Some philosophers believe that such a derivation is forever impossible because of the very nature of the two fields. But the proofs attempted so far for this thesis are certainly insufficient (1938, p. 60).

Later he has the same sort of thing to say about the reduction of psychology or other social sciences to biology.

> A different and less linguistic approach is to contrast the unity of scientific subject matter with the unity of scientific method. Many would agree that different sciences have different subject matters; for example, in no real sense is the subject matter of astronomy the same as that of psychopharmacology. But many would affirm that in spite of the radically different subject matters of science there are important ways in which the methods of science are the same in every domain of investigation. The most obvious and simple examples immediately come to mind. There is not one arithmetic for psychological theories of motivation and another for cosmological theories of the universe. More generally, there are not different theories of the differential and integral calculus or of partial differential equations or of probability theory.

There is a great mass of mathematical methods and results that are available for use in all domains of science and that are, in fact, quite widely used in very different parts of science. There is a plausible prima facie case for the unity of science in terms of unity of scientific method. This may be one of the most reasonable meanings to be attached to any central thesis about the unity of science. However, I shall be negative even about this thesis in the sequel.

2. UNITY AND REDUCTIONISM

What I have said earlier about different sciences having obviously different subject matters was said too hastily because there is a historically important sense of unity. One form or another of reductionism has been central to the discussion of unity of science for a very long time. I concentrate on three such forms: reduction of language, reduction of subject matter, and reduction of method.

Reduction of language. Carnap's views about the reduction of the language of science to commonsense language about physical objects remain appealing. He states his general thesis in such a way that no strong claims about the reduction of psychology to physics, for example, are implied, and I am sure much is correct about what he has had to say. On the other hand, it seems appropriate to emphasize the very clear senses in which there is no reduction of language. The reduction certainly does not take place in practice, and it may be rightly claimed that the reduction in theory remains in a hopelessly vague state.

There are many ways to illustrate the basis for my skepticism about any serious reduction of language. Part of my thesis about the plurality of science is that the languages of the different branches of science are diverging rather than converging as they become increasingly technical. Let me begin with a personal example. My daughter Patricia is taking a PhD in neurophysiology, and she recently gave me a subscription to what is supposed to be an expository journal, entitled *Neurosciences: Research Program Bulletin.* After several efforts at reading this journal, I have reached the conclusion that the exposition is only for those in nearby disciplines. I quote one passage from an issue (1976) dealing with neuron-target cell interactions.

> The above studies define the anterograde transsynaptic regulation of adrenergic ontogeny. Black and co-workers (1972b) have also demonstrated that postsynaptic neurons regulate presynaptic development through a retrograde process. During the course of maturation, presynaptic ChAc activity increased 30- to 40-fold (Figure 19), and this rise paralleled the formation of ganglionic synapses (Figure 20). If postsynaptic adrenergic neurons in neonatal rats were chemically destroyed with 6-hydroxydopamine (Figure 24) or immunologically destroyed with antiserum to NGF (Figure 25), the normal development of presynaptic ChAc activity was prevented. These data, viewed in conjunction with the anterograde regulation studies lead to the conclusion that there is a bidirectional flow of regulatory information at the synapses during development (1976, p. 253).

This is by no means the least intelligible passage. It seems to me it illustrates the cognitive facts of life. The sciences are diverging and there is no reason to think that any kind of convergence will ever occur. Moreover, this divergence is not something of recent origin. It has been present for a long time in that oldest of quantitative sciences, astronomy, and it is now increasingly present throughout all branches of science.

There is another point I want to raise in opposition to a claim made by some philosophers and philosophically minded physicists. Some persons have held that in the physical sciences at least, substantial theoretical unification can be expected in the future and, with this unification, a unification of the theoretical language of the physical sciences, thereby simplifying the cognitive problem of understanding various domains. I have skepticism about this thesis that I shall explain later, but at this point I wish to emphasize that it takes care of only a small part of the difficulties. It is the experimental language of the physical sciences as well as of the other sciences that is difficult to understand, much more so for the outsider than the theoretical language. There is, I believe, no comparison in the cognitive difficulty for a philosopher of reading theoretical articles in quantum mechanics and reading current experimental articles in any developed branch of physics. The experimental literature is simply impossible to penetrate without a major learning effort. There are reasons for this impenetrability that I shall not attempt to go into on this occasion but stipulate to let stand as a fact.

Personally I applaud the divergence of language in science and find in it no grounds for skepticism or pessimism about the continued growth of science. The irreducible pluralism of languages of science is as desirable a feature as is the irreducible plurality of political views in a democracy.

Reduction of subject matter. At least since the time of Democritus in the 5th century B.C., strong and attractive theses about the reduction of all phenomena to atoms in motion have been set forth. Because of the striking scientific successes of the atomic theory of matter since the beginning of the 19th century, this theory has dominated the views of plain men and philosophers alike. In one sense, it is difficult to deny that everything in the universe is nothing but some particular swarm of particles. Of course, as we move into the latter part of the 20th century, we recognize this fantasy for what it is. We are no longer clear about what we mean by particles or even if the concept as originally stated is anywhere near the mark. The universe is indeed made of something but we are vastly ignorant of what that something is. The more we probe, the more it seems that the kind of simple and orderly view advanced as part of ancient atomism and that seemed so near realization toward the end of the 19th century is ever further from being a true description. To reverse the phrase used earlier, it is not swarms of particles that things are made of, but particles that are made of swarms. There are still physicists about who hold that we will one day find the ultimate simples out of which all other things are made, but as such claims have been continually revised and as the complexity of high-energy physics and elementary particle theory has increased, there seems little reason that we shall ever again be able to seriously believe in the strong sense of reduction that Democritus had attractively formulated.

To put the matter in a skeptical fashion, we cannot have a reduction of subject matter to the ultimate physical entities because we do not know what those entities are. I have on another occasion (1974a) expressed my reasons for holding that Aristotle's theory of matter may be sounder and more sensible than the kind of simpleminded atomistic reductionist views dominating our thinking about the physical world for 200 years.

There is another appealing argument against reduction of subject matter in the physical sense that does not rest on the controversy about the status of mental events but on what has happened in the development of computers. Perhaps for the first time we have become fully and completely aware that the same cognitive structures can be realized in physically radically different ways. I have in mind the fact that we now have computers that are built on quite different physical principles for example, old computers using vacuum tubes and modern computers using semiconductors can execute exactly the same programs and can perform exactly the same tasks. The differences in physical properties are striking between these two generations of computers. They stand in sharp contrast to different generations of animal species, which have very similar physical constitutions but which may have very different cultural histories. It has often been remarked upon that men of quite similar constitutions can have quite different thoughts. The computer case stands this argument on its head—it is not that the hardware is the same and the software different but rather that the hardware is radically different and the software of thoughts the same. Reduction in this situation, below the level of the concepts of information processing, seems wholly uninteresting and barren. Reduction to physical concepts is not only impractical but also theoretically empty.

The same kinds of arguments against reductionism of subject matter can be found even within physics. A familiar example is the currently accepted view that it is hopeless to try to solve the problems of quantum chemistry by applying the fundamental laws of quantum mechanics. It is hopeless in the same way that it is hopeless to program a computer to play the perfect chess game by always looking ahead to all possible future moves. The combinatorial explosion is so drastic and so overwhelming that theoretical arguments can be given that not only now but also in the future it will be impossible by direct computation to reduce the problems of quantum chemistry to problems of ordinary quantum mechanics. Quantum chemistry, in spite of its proximity to quantum mechanics, is and will remain an essentially autonomous discipline. At the level of computability, reduction is not only practically impossible but theoretically so as well.

An impressive substantive example of reduction is the reduction of large parts of mathematics to set theory. But even here, the reduction to a single subject matter of different parts of mathematics has a kind of barren formality about it. It is not that the fact of the reduction is conceptually uninteresting but rather that it has limited interest and does not say much about many aspects of mathematics. Mathematics, like science, is made up of many different subdisciplines, each going its own way and each primarily sensitive to the nuances of its own subject matter. Moreover, as we have reached for a deeper understanding of the foundations of mathematics we have come to realize that the foundations are not to be built on a bedrock of certainty but that, in many ways developed parts of mathematics are much better understood than the foundations themselves. As in the case of physics, an effort of reduction is now an effort of reduction to we know not what.

In many ways a more significant mathematical example is the reduction of computational mathematics to computability by Turing machines, but as in the case of set theory, the reduction is irrelevant to most computational problems of theoretical or practical interest.

Reduction of method. As I remarked earlier, many philosophers and scientists would claim that there is an important sense in which the methods of science are the same in every domain of investigation. Some aspects of this sense of unity, as I also noted, are well recognized and indisputable. The common use of elementary mathematics and the common teaching of elementary mathematical methods for application in all domains of science can scarcely be denied. But it seems to me it is now important to emphasize the plurality of methods and the vast difference in methodology of different parts of science. The use of elementary mathematics—and I emphasize *elementary* because almost all applications of mathematics in science are elementary from a mathematical standpoint—as well as the use of certain elementary statistical methods does not go very far toward characterizing the methodology of any particular branch of science. As I have emphasized earlier, it is especially the experimental methods of different branches of science that have radically different form. It is no exaggeration to say that the handbooks of experimental method for one discipline are generally unreadable by experts in another discipline (the definition of 'discipline' can here be quite narrow). Physicists working in solid-state physics cannot intelligibly read the detailed accounts of method in other parts of physics. This is true even of less developed sciences like psychology. Physiological psychologists use a set of experimental methods that are foreign to psychologists specializing, for example, in educational test theory, and correspondingly the intricate details of the methodology of test construction will be unknown to almost any physiological psychologist.

Even within the narrow domain of statistical methods, different disciplines have different statistical approaches to their particular subject matters. The statistical tools of psychologists are in general quite different from those of economists. Moreover, within a single broad discipline like physics, there are in different areas great variations in the use of statistical methods, a fact that has been well documented by Paul Humphreys (1976).

The unity of science arose to a fair degree as a rallying cry of philosophers trying to overcome the heavy weight of 19th-century German idealism. A half century later the picture looks very different. The period since the Encyclopedia of Unified Science first appeared has been the era of greatest development and expansion of science in the history of thought. The massive enterprise of science no longer needs any philosophical shoring up to protect it from errant philosophical views. The rallying cry of unity followed by three cheers for reductionism should now be replaced by a patient examination of the many ways in which different sciences differ in language, subject matter, and method as well as by synoptic views of the ways in which they are alike. Related to unity and reduction are the two long-standing themes of certainty of knowledge and completeness of science. In making my case for the plurality of science, I want to say something about both of these unsupported dogmas.

3. THE SEARCH FOR CERTAINTY

From Descartes to Russell, a central theme of modern philosophy has been the setting forth of methods by which certainty of knowledge can be achieved. The repeatedly stated intention has been to find a basis that is, on the one hand, certain and, on the other hand, adequate for the remaining superstructure of knowledge, including science. The introduction of the concept of sense data and the history of the use of this concept have dominated the search for certainty in knowledge, especially in the empirical tradition, as an alternative to direct rational knowledge of the universe.

All of us can applaud the criticism of rationalism and the justifiable concern not to accept the possibility of direct knowledge of the world without experience. But it was clearly in a desire to compete with the kind of foundation that rationalism offered that the mistaken additional step was taken of attempting to ground knowledge and experience in a way that guaranteed certainty for the results. The reduction of the analysis of experience to sense data is itself one of the grand and futile themes of reductionism, in this case largely driven by the quest for certainty. Although it is not appropriate to pursue the larger epistemological issues involved, I would like to consider some particular issues of certainty that have been important in the development of modern scientific methods.

Errors of measurement. With the development of scientific methodology and probability theory in the 18th century, it was recognized that not only did errors in measurements rise but also that a systematic theory of these errors could be given. Fundamental memoirs on the subject were written by Simpson, Lagrange, Laplace, and others. For our purposes, what is important about these memoirs is that there was no examination of the question of the existence or nonexistence of an exact value for the quantity being measured. It was implicit in these 18th-century developments, as it was implicit in Laplace's entire theory of probability, that probabilistic considerations, including errors, arise from ignorance of true causes and that the physical universe is so constituted that in principle we should be able to achieve the exact true value of any measurable physical quantities. Throughout the 19th century it was implicit that it was simply a matter of tedious and time-consuming effort to refine the measured values of any quantity one more significant digit. Nothing fundamental stood in the way of making such a refinement. It is a curious and conceptually interesting fact that, as far as I know, no one in this period enunciated the thesis that this was all a mistake, that there were continual random fluctuations in all continuous real quantities, and that the concept of an exact value had no clear meaning.

The development of quantum mechanics in this century made physicists reluctantly but conclusively recognize that it did not make sense to claim that any physical quantity could be measured with arbitrary precision in conjunction with the simultaneous measurement of other related physical quantities. It was recognized that the inability to make exact measurement is not due to technological inadequacies of measuring equipment but is central to the fundamental theory itself.

Even within the framework of quantum mechanics, however, there has tended to be a large conceptual equivocation on the nature of uncertainty. On the one hand, the claim has been that interference from the measuring apparatus makes uncertainty a necessary consequence. In this context some aspects of uncertainty need to be noted. It is not surprising that if we measure human beings at different times and places we expect to get different measurements of height and weight. But in the case of quantum mechanics what is surprising is that variation is found in particles submitted to "identical" experimental preparations. Once again a thesis of simplicity and unity is at work. Electrons should differ only in numerical identity, not in any of their properties. And if this is not true of electrons, there should be finer particles discoverable that do satisfy such a principle of identity.

The other view, and the sounder one in my judgment, is that random fluctuations are an intrinsic part of the behavior of microscopic phenomena. No process of measurement is needed to generate these fluctuations; they are a part of nature and lead to a natural view of the impossibility of obtaining results of arbitrary precision about microscopic physical quantities.

If we examine the status of theory and experiments in other domains of science, it seems to me that similar claims about the absence of certainty can be made. The thrust for certainty associated with classical physics, British empiricism, and Kantian idealism is now spent.

4. THE SEARCH FOR COMPLETENESS

Views about the unity of science, coupled with views about the reduction of knowledge to an epistemologically certain basis like that of sense data, are often accompanied by an implicit doctrine of completeness. Such a doctrine is often expressed by assumptions about the uniformity of nature and assumptions about the universe being ultimately totally ordered and consequently fully knowable in character. Unity, certainty, and completeness can easily be put together to produce a delightful philosophical fantasy.

In considering problems of completeness, I begin with logic and mathematics but have as my main focus the subsequent discussion of the empirical sciences.

Logic is the one area of experience in which a really satisfactory theory of completeness has been developed. The facts are too familiar to require a detailed review. The fundamental result is Gödel's completeness theorem that in first-order logic a formula is universally valid if and only if it is logically probable. Thus, our apparatus of logical derivation is adequate to the task of deriving any valid logical formula, that is, any logical truth. What we have in first-order logic is a happy match of syntax and semantics.

On the other hand, as Kreisel has emphasized in numerous publications (e.g., (1967)), this match of syntax and semantics is not used in the proof of logical theorems. Rather, general set-theoretical and topological methods are continually drawn upon. One reason is that proofs given in the syntax of elementary logic are psychologically opaque and therefore in nontrivial cases easily subject to error. Another is that it is not a natural setting for studying the relation of objects that are the focus of the theory to other related objects; as an example, even the numerical representation theorem for simple orderings cannot be proved in first-order fashion. Completeness of elementary logic is of some conceptual interest, but from a practical mathematical standpoint useless.

Incompleteness of arithmetic. The most famous incompleteness result occurs at an elementary level, namely, at the level of arithmetic or elementary number theory. In broad conceptual terms, Gödel's result shows that any formal system whose language is rich enough to represent a minimum of arithmetic is incomplete. A much earlier and historically important incompleteness result was the following.

Incompleteness of geometric constructions. The three classical construction problems that the ancient Greeks could not solve by elementary means were those of trisecting an angle, doubling a cube, and squaring a circle. It was not until the 19th century that these constructions were shown to be impossible by elementary means, thereby establishing a conceptually important incompleteness result for elementary geometry.

Incompleteness of set theory. In the latter part of the 19th century, on the basis of the work of Frege in one direction and Cantor in another, it seemed that the theory of sets or classes was the natural framework within which to construct the rest of mathematics. Research in the 20th century on the foundations of set theory, some of it recent, has shown that there is a disturbing sense of incompleteness in set theory, when formulated as a first-order theory. The continuum hypothesis as well as the axiom of choice is independent of other principles of set theory, and, as in the case of geometry, a variety of set theories can be constructed, at least first-order set theories.

The continuum hypothesis, for example, is decidable in second-order set theory, but we do not yet know in which way, that is, as true or false. Thus there is clearly less freedom for variation in second-order set theory, but also at present much less clarity about its structure. The results of these various investigations show unequivocally that the hope for some simple and complete foundation of mathematics is not likely to be attained.

Theories with standard formalization. The modern logical sense of completeness for theories with standard formalization, that is, theories formalized within first-order logic, provides a sharp and definite concept that did not exist in the past. Recall that the characterization of completeness in this context is that a theory is complete if and only if every sentence of the theory is either valid in the theory or inconsistent with the theory—that is, its negation is valid in the theory.

In back of this well-defined logical notion is a long history of discussions in physics that are vaguer and less sharply formulated but that have a similar intuitive content.

Kant's sense of completeness. Although there is no time here to examine this history, it is worth mentioning the high point of its expression as found in Kant's *Metaphysical Foundations of Natural Science.* Kant's claim is not for the completeness of physics but for the completeness of the metaphysical foundations of physics. After giving the reason that it is desirable to separate heterogeneous principles in order to locate errors and confusions, he gives as the second reason the argument concerning completeness.

> There may serve as a second ground for recommending this procedure the fact that in all that is called metaphysics the absolute completeness of the sciences may be hoped for, which is of such a sort as can be promised in no other kind of cognitions; and therefore just as in the metaphysics of nature in general, so here also the completeness of the metaphysics of corporeal nature may be confidently expected....
>
> The schema for the completeness of a metaphysical system, whether of nature in general or of corporeal nature in particular, is the table of the categories. For there are no more pure concepts of the understanding, which can concern the nature of things. (1970, pp. 10–11).

It need scarcely be said that Kant's argument in terms of the table of the categories scarcely satisfied 18th-century mathematical standards, let alone modern ones. His argument for completeness was not subtle, but his explicit focus on the issue of completeness was important and original.

The unified field theory. After Kant, there was important system building in physics during the 19th century, and there were attempts by Kelvin, Maxwell, and others to reduce all known physical phenomena to mechanical models, but these attempts were not as imperialistic and forthright in spirit as Kant's. A case can be made, I think, for taking Einstein's general theory of relativity, especially the attempt at a unified field theory, as the real successor to Kant in the attempt to obtain completeness. I do not want to make the parallel between Kant and Einstein too close, however, for Einstein does not hold an a priori metaphysical view of the foundations of physics. What they do share is a strong search for completeness of theory. Einstein's goal was to find a unified field theory defining one common structure from which all forces of nature could be derived. In the grand version of the scheme, for given boundary conditions, the differential equations would have a unique solution for the entire universe, and all physical phenomena would be encompassed within the theory. The geometrodynamics of John Wheeler and his collaborators is the most recent version of the Einstein vision. Wheeler, especially, formulates the problem in a way that is reminiscent of Descartes: "Are fields and particles foreign entitles immersed *in* geometry, or are they nothing *but* geometry?" (1962, p. 361).

Had the program of Einstein and the later program of Wheeler been carried to completion, my advocacy of skepticism toward the problem of completeness in empirical science would have to retreat from bold assertion of inevitable incompleteness. However, it seems to me that there is, at least in the current scientific temperment, total support for the thesis of incompleteness. Grand building of theories has currently gone out of fashion in fields as far apart as physics and sociology, and there seems to be a deeper appreciation of the problems of ever settling, in any definitive way, the fundamental laws of complex phenomena.

As the examples I have mentioned—and many others that I have not—demonstrate, in most areas of knowledge it is too much to expect theories to have a strong form of completeness. What we have learned to live with in practice is an appropriate form of completeness, but we have not built this working practice explicitly into our philosophy as thoroughly as we might. It is apparent from various examples that weak forms of completeness may be expected for theories about restricted areas of experience. It seems wholly inappropriate, unlikely, and, in many ways, absurd to expect theories that cover large areas of experience, or, in the most grandiose cases, *all* of experience, to have a strong degree of completeness.

The application of working scientific theories to particular areas of experience is almost always schematic and highly approximate in character. Whether we are predicting the behavior of elementary particles, the weather, or international trade—any phenomenon, in fact, that has a reasonable degree of complexity—we can hope only to encompass a restricted part of the phenomenon.

It is sometimes said that it is exactly the role of experimentation to isolate particular fragments of experience that can be dealt with in relatively complete fashion. This is, I think, more a dogma of philosophers who have not engaged in much experimentation than it is of practicing experimental scientists. When involved in experimentation, I have been struck by how much my schematic views of theories also apply to experimental work. First one concrete thing and then another is abstracted and simplified to make the data fit within the limited set of concepts of the theory being tested.[1]

Let me put the matter another way. A common philosophical conception of science is that it is an ever closer approximation to a set of eternal truths that hold always and everywhere. Such a conception of science can be traced from Plato through Aristotle and onward to Descartes, Kant, and more recent philosophers, and this account has no doubt been accepted by many scientists as well. It is my own view that a much better case can be made for the kind of instrumental conception of general terms by Peirce, Dewey, and their successors. In this view scientific activity is perpetual problem solving. No area of experience is totally and completely settled by providing a set of basic truths; but rather, we are continually confronted with new situations and new problems, and we bring to these problems and situations a potpourri of scientific methods, techniques, and concepts, which in many cases we have learned to use with great facility.

The concept of objective truth does not directly disappear in such a view of science, but what we might call the cosmological or global view of truth is looked at with skepticism just as is a global or cosmological view of completeness. Like our own lives and endeavors, scientific theories are local and are designed to meet a given set of problems. As new problems arise new theories are needed, and in almost all cases the theories used for the old set of problems have not been tested to the fullest extent feasible nor been confirmed as broadly or as deeply as possible, but the time is ripe for something new, and we move on to something else. Again this conception of science does not mean that there cannot be continued correction in a sequence of theories meeting a particular sequence of problems; but it does urge that the sequence does not necessarily converge. In fact, to express the kind of incompleteness I am after, we can even make the strong assumption that in many domains of experience the scientific theory that replaces the best old theory is always an improvement, and therefore we have a kind of monotone increasing sequence. Nonetheless, as in the case of a strictly monotone increasing sequence of integers, there is no convergence to a finite value—the sequence is never completed—and so it is with scientific theories. There is no bounded fixed result toward which we are converging or that we can hope ever to achieve. Scientific knowledge, like the rest of our knowledge, will forever remain pluralistic and highly schematic in character.

[1] This idea is developed in some detail in Suppes (1962).

5

HEURISTICS AND THE AXIOMATIC METHOD

1. THE PLACE OF THE AXIOMATIC METHOD

Over the last 100 years a variety of arguments have been given for using the axiomatic method in mathematics and in science. There is not uniform agreement that the method is always appropriate or useful. However, there is, I think, general agreement that the use of such methods has revolutionized the presentation of mathematics and has had significant impact in the empirical sciences as well.

Various arguments in favor of giving an explicit axiomatic analysis of structures in a given discipline have been given. The standard arguments concentrate on matters of clarity, explicitness, generality, objectivity, and self-containedness (Suppes, 1968).

Arguments of another sort are sometimes found in physics. I quote one example from quantum field theory (Bogolubov, Logunov, Todorov, 1975).

> It is widely believed that axiomatization is a kind of polishing, which is applied to an area of science after it has been, for all practical purposes, completed. This is not true, even

*Reprinted from *Methods of heuristics* (ed. by R. Groner, M. Groner and W. F. Bischof), 1983, pp. 79–88. Hillsdale, NJ: Erlbaum. I am indebted to Georg Kreisel for several useful criticisms of an earlier draft.

> in pure mathematics. Admittedly, the modern axiomatization of arithmetic and Euclidean geometry marked the completion of these disciplines (although at the same time it stimulated a new science—mathematical logic, or metamathematics). For most areas of contemporary mathematics, however, such as functional analysis, axiomatization is a fundamental method of exploration, a starting point. (Of course, the system of axioms may be modified as the subject develops.) In theoretical physics, since the time of Newton, the axiomatic method has served not only for the systematization of results previously obtained, but also in the discovery of new results (p. 1).

What I want to do in the present chapter is rather similar. The argument I want to concentrate on is sometimes stated very informally, but it is often implicit and behind the scenes. It is that the use of the axiomatic method has a positive heuristic value in understanding a subject, in solving problems in it, and in formulating new problems. At the most satisfactory level, this chapter would contain some conceptual ideas about the heuristic value of the axiomatic method and would then go on to present detailed empirical evidence in support of or against these conceptual claims. As you might imagine, I am not able to provide any detailed empirical data, but what I have to say should in principle be testable.

I also want to make clear that my analysis is not meant to be a panegyric for the axiomatic method. Application of the method in some parts of science has had a negative effect. I should also mention that, in spite of the fact that axiomatic methods have certainly been developed and applied mainly in pure mathematics, I consider on an equal basis the physical and social sciences.

When axioms are appropriate. The preceding quotation from a well-known treatise on quantum field theory represents one important viewpoint. I now want to move to the social sciences. Economics uses mathematical methods, and in particular axiomatic methods, to a much greater extent than any other social science. As some of my economist friends put it, you have to know some modern mathematics in order not to become technologically obsolete as an economist. In a subject like economics that has been developed over many years, that has close ties to politics in many of its intellectual roots and that often reflects strong national biases, the virtues of the extensive use of mathematical methods, and especially axiomatic methods, are apparent. Economists from all parts of the world converse easily and clearly about their basic assumptions when they operate within the axiomatic frameworks that are so much a common part of contemporary work in economics.

For different reasons the use of axiomatic methods has played a similar positive role in sociology. For over 100 years sociology has been almost overwhelmed with large-scale, high-sounding theories. Inside some of these theories are some interesting and original ideas, but these creative seeds have often been lost in the mounds of chaff. What is happening in sociology reminds me of what has sometimes been said of the 17th century; Newton's *Principia* was not only a work marvelous for its deep results but also for its intellectual purity and austerity. The speculative and overwrought ideas of Descartes and others about the nature of matter and of physical phenomena, exemplified most strikingly in Descartes' *Principles of Philosophy,* were replaced by something that was substantial and solid throughout. The modern tendency in sociology represents a corresponding move from Cartesian method to Newtonian analysis. Good examples of modern work are Coleman (1964), Fararo (1973), and Blalock, Aganbegian, Borodkin, Boudon, and Capecchi (1975). These three works are of quite a different sort. Coleman's treatise is an early and influential book in the extended application of mathematical methods to standard problems in sociology. Fararo's book is closer to being a standard textbook, and the multiple-authored book edited by Blalock et al. provides reprints of many current articles relevant to mathematical and axiomatic studies in sociology. The material in these three volumes is a far cry from the kind of philosophical sociology that is still very prominent in many parts of Europe and that was dominant throughout the world a few decades ago. The mathematical methods in sociology I am referring to have the heuristic virtue of forcing those who use them to achieve a certain degree of explicitness and precision of formulation. It is too easily forgotten how important it is to convert certain subject matters from vague qualitative discussions to disciplined mathematically based discourse. It is also too easy to think that this is a problem that has only been faced by the social sciences. A little reading in Descartes, Boscovich, or any of a number of other authors provides evidence that physics had a similar problem before the 19th century.

The story is somewhat different in psychology, which has always been more data bound and experimentally bound than either economics or sociology. The earlier attempts at axiomatization in psychology were more in the spirit of Descartes than Newton. Leibniz said in a famous phrase that Descartes' treatise on physics, the *Principles of Philosophy* just mentioned, was a *roman de physique.* I have said the same of Piaget's attempts at axiomatization in psychology and have called them a *roman de psychologie* (Suppes, 1973a). I would say the same also of the earlier work of the learning theorists of the 1930s and 40s, for example, Tolman and Hull. Do not misunderstand me, novels are not necessarily bad; they have a place even in science. The speculative system of Descartes played in its own way a major role in the development of physics in the 17th century. The same can be said, in an even more positive way, about the work of Piaget, Tolman, and Hull.

Since 1950 there has been a great variety of axiomatic work in psychology, most of it closely linked to experimental data. I am thinking of the work in learning theory, decision theory, measurement theory, formal models of perception, and psychophysical processes, to mention what are perhaps the most important areas. On the other hand, the story has not been one of unmitigated success. Much of the work in contemporary cognitive psychology is not even mathematical in character, let alone axiomatic. There is no doubt a feeling among many cognitive psychologists that it is premature to think of the development of cognitive structures in mathematical terms. I do not think these cognitive psychologists holding the views I attribute to them are entirely wrong. They are just misguided!

2. HEURISTIC VERSUS NONHEURISTIC AXIOMS

I assume for the remainder of this chapter that for many scientific theories it is appropriate to attempt to give a thorough axiomatic treatment. What I want to do is to classify various axiomatic analyses as heuristic or not. By an axiomatic analysis being "heuristic," I mean that the analysis yields axioms that seem intuitively to organize and facilitate our thinking about the subject, and in particular our ability to formulate, in an ordinarily self-contained way, problems concerned with the phenomena governed by the theory and their solutions.

In considering these examples, I have in mind that the axiomatic method is relatively neutral regarding its heuristic value. It seems to me that there are examples, well-known in fact in the literature, that do facilitate our thinking. On the other hand, there are also well-known examples that represent a sophisticated mathematical foundation of a discipline, but that are formulated in such a way that they prohibit natural and intuitive ways of thinking about problems, especially new problems in the discipline. By calling some axiomatic analyses unheuristic, I do not mean to suggest that they do not have value for other reasons. I do mean to suggest that they do not represent the kind of transparent and conceptually satisfactory solution we should aim at whenever possible.

First heuristic example: field of real numbers. The construction of real numbers by Dedekind cuts or as equivalence classes of Cauchy sequences completes an important 19th-century program on the arithmetization of analysis, but the resulting objects, taken as the real numbers in a literal fashion, are unnatural to deal with. In contrast, the standard axioms for the field of real numbers using the least upper-bound axiom for completeness seem very natural and intuitive. The axioms express the algebraic content of the rational operations on the real numbers in a simple and elegant way (I recognize, of course, that there is a slight variance on how the axioms are formulated in this respect, but these minor variations are not of concern here). Moreover, the least upper-bound axiom, to the effect that every nonempty bounded set of real numbers has a least upper bound, also seems easy to comprehend, even though it has a very different character from the other axioms. Elementary proofs in real analysis can use these axioms in a way that is easy for students to understand and for instructors to explain. Part of the heuristic value of the axioms is, I believe, that all but the least upper-bound axiom can be formulated with free variables only. This leaves the algebraic structure transparent and easy for the student to manipulate.

Second heuristic example: Kolmogorov's axioms for probability. To appreciate the clarity and definite intuitive foundation Kolmogorov (1933) gave to the concept of probability in his well-known axiomatization, one needs only to examine the literature on the foundations of probability prior to his work. Even basic general properties were not entirely clear. Certainly the appropriate generality was not obtained together with axioms whose conceptual foundation was easy to understand. By formulating the axioms in terms of a measure on a algebra of sets, with the sets interpreted as events, he provided an axiomatic foundation that has dominated 50 years of probability theory. The earlier work of Borel and Keynes, for example, lacked both clarity and generality. Also important in Kolmogorov's treatment was the explicit introduction of random variables as the main tool used in advanced probability work.

To a remarkable degree, Kolmogorov's approach has simply obliterated in the mathematical literature of probability theory the earlier foundational formulations. Before Kolmogorov's work it used to be said that probability was a subject that could not be treated in a proper mathematical fashion because the foundations were so unclear. The heuristic value of Kolmogorov's work was to clear away the underbrush of the past and leave a new and adequate axiomatic formulation standing unsupported by any need for historical references to earlier work. This elimination of the past is one of the great heuristic virtues of simplification that the axiomatic method can achieve when used in the best possible form.

It is also important to recognize that a brilliant piece of axiomatic work like that of Kolmogorov need not be in any absolute sense final. It is just that it provides a basis for going forward in a new and unencumbered fashion.

For many reasons, I do not think that Kolmogorov's axioms are really the natural ones for many physical applications, but this is a minor complaint in the perspective of what was accomplished by his axiomatic presentation in the 1930s.

A nonheuristic example: Mackey's axioms for quantum mechanics. Because of its mathematical clarity and thoroughness, Mackey's (1957, 1963) axiomatic foundations of classical quantum mechanics have been generally well-received and cited often as the standard work on the subject. I take the view here that heuristically this is a bad example of axiomatization. As might be expected, I hope that what I have to say will be intrinsically more interesting than the rather laudatory general things I said about the two previous examples cited as good heuristic instances.

Let me first try to put in a general way my central objection to Mackey's axiomatization. There are two main points I want to make. First, the axioms about the probability distribution of operators are formulated for single operators. There is no natural discussion about the causal development of a quantum-mechanical system and, consequently, the way in which one would intuitively think of a temporal sequence of operators being causally related. I expect, of course, that these causal relations will be stochastic in nature, but they are intuitively important to consider, indeed essential to the dynamical aspects of the theory. Second, if we think in natural terms of the trajectory of a particle, for example, we must think of it as a continuous sequence of operators being able to ask at each instant in time in which Borel sets the value of the operator lies. I submit that if physics had started this way, no serious complex problem would ever have been solved. A more natural and intuitive way of thinking of trajectories of particles is needed. It might be said that Mackey is just cleaning up what the physicists have said in an informal way. I think that a case can be made for this. My point is not to criticize Mackey's work as introducing discrepancies between the way physicists talk and the axioms he has given, but rather that the axioms taken literally present a wrong picture of how to think about physical problems in quantum mechanics.

I mention at this point the more important of Mackey's axioms. Briefly speaking, Mackey proceeds in the following fashion for the time-independent case. Let Θ be the set of observables and let S be the set of states; any structure on the sets Θ and S is explicitly stated in the axioms. The function $p(A, \alpha, E)$ is defined whenever $A \in \Theta, \alpha \in S$ and E is a Borel set of real numbers. Intuitively $p(A, \alpha, E)$ is the probability of measuring observable A in set E when the state of the system is α. The first axiom states in fact that for every A in Θ and α in S, $p(A, \alpha, E)$ is a probability measure in the argument E on the set of all real numbers. The second axiom guarantees uniqueness of observables with a given probability distribution, and similarly for states. It is a kind of extensionality axiom for observables and states.

If $p(A, \alpha, E) = p(A', \alpha, E)$ *for all* α *in* S *and Borel sets* E *then* $A = A'$, *and if* $p(A, \alpha, E) = p(A, \alpha', E)$ *for all* A *in* Θ *and all Borel sets* E *then* $\alpha = \alpha'$.

The remaining axioms are more technical and are not given here. Properties as two-valued observables are defined, and a certain partial ordering in terms of probability distributions on properties is defined. The final and most powerful axiom is then the assertion that the set of all properties under the given ordering is isomorphic to the partially ordered set of all closed subspaces of a separable infinite-dimensional complex Hilbert space.

The last axiom also makes clear another heuristic weakness. The correspondence between operators and observables is left at the postulation of a one-to-one correspondence. Clearly, not much real physics could be done within this framework. What is important from the standpoint of physics is the derivation of the important correspondences and the provision of tools for the derivation of others that may be wanted. Thus, the various arguments that are given for the standard operator for position and the standard operator for momentum need to be, I would claim, incorporated directly into the axiomatic framework in order to have a heuristically acceptable set of axioms.

In criticizing so severely Mackey's axioms from a heuristic standpoint, I am not suggesting that it is either obvious or easy how to replace them by axioms for quantum mechanics that are heuristically of the right sort. Mackey (1963) himself agrees with this point: "It is not yet possible to deduce the present form of quantum mechanics from completely plausible and natural axioms (p. 62)." My view is that, given the way in which classical quantum mechanics developed historically, only a rather radical shift in our thinking will lead to a heuristically transparent formulation.

Starting with a quotation from the treatise of Bogolubov et al., on quantum field theory, I have stressed the heuristic value of the axiomatic method in simplifying subject matters so as to make discoveries easier and the exposition of subjects pedagogically more accessible. It seems to me that Mackey's treatment fails on both these points in spite of the other virtues of his classical work.

3. AXIOMATIC ANALYSIS OF QUALITATIVE DERIVATIONS

It is widespread folklore in physics and engineering that if one gets diverted into proving theorems nothing of real interest will come out from a physical standpoint. The focus of physics should be on solving problems and not on proving theorems. Leave that business to the mathematicians. We could agree that there is a proper division of work here and that, moreover, without any consideration of division of work it is still better and more useful to have the solution of a hard problem than the proof of a trivial theorem (or vice versa).

It is also a standard claim that graduate students in physics and in engineering are not expected to be able to prove theorems. It is not a part of their training. When physicists discuss axiomatics, it is sometimes called *physical* axiomatics, because it is not meant to have the status of a genuine axiomatic analysis.

There is little doubt about the importance attached in engineering and physics to students' having the ability to derive from qualitative principles an appropriate differential equation. Such derivations are considered an essential part of problem-solving skills. Good teachers have a lot of important things to say about how one is to think about such derivations, but the systematic theory is quite undeveloped. Indeed, it is a common thing to juxtapose the quite informal state of such problem solving to formal theorem proving. However, I think this is a false division. We should be able to give an axiomatic analysis of such qualitative derivations in the same spirit that we analyze other systematic phenomena. Now it is quite true that this axiomatic analysis could miss the heuristic spirit that seems so central to learning how to make such derivations, but it should be an important criterion of evaluation that the axioms do not miss this heuristic spirit.

I want to be clear that the giving of such an axiomatic analysis of the foundation of qualitative derivations in physics, engineering, and other sciences does not in itself constitute a heuristic analysis, but I think that it is an important and essential step that will guide students in their attempts to give proper derivations.

What I am asking for corresponds in many ways to the analysis of the concept of mathematical proof, but I am not interested here in the direction so characteristic of proof theory, namely, the reduction of proofs to an explicit form consisting of a large number of elementary steps. I am interested more in the analysis of mathematical proofs as they are presented by good writers in textbooks and treatises. This later kind of analysis, which is what many mathematicians expect, I think, from proof theory, is nearly as undeveloped as the analysis of qualitative derivations.

Mathematicians are generally aware of the rigorous and explicit theory of proofs that has developed since Hilbert, but they are not always sensitive to the differences between this explicit formal theory and the actual practices in giving proofs in mathematics. As we move from mathematics to sciences, awareness of this gap assumes a different form. In the classical tradition of problem solving I have alluded to, there is no concern at all for formal proofs. In spite of these differences, I think there is a natural and homogeneous common ground that can be occupied by a theory of informal mathematical proofs and a theory of qualitative derivations in engineering, physics, and the other sciences.

There is a final historical point I want to make on the place of the axiomatic method in the kind of analyses I have just been discussing. It is commonly observed that the explicit role of the axiomatic method in mathematics is less important now than it was at the turn of the century. The detailed discussion of axioms and consideration of their independence, consistency, and completeness were the focus of intensive inquiry especially in the foundations of geometry. The lack of current concern for such questions can be seen in the lack of attention they receive in the treatise of Bourbaki covering so many parts of modern mathematics. As developments of particular disciplines have matured, the emphasis has shifted from explicit axiomatic methods to the identification of basic structures. It is this identification of basic structures without regard to the finer points of the axiomatic assumptions that is characteristic of Bourbaki and many other systematic treatises in modern mathematics.

It seems to me, on the other hand, that the rather primitive status of the theory of qualitative derivations or the theory of informal proofs is precisely a subject calling for a sustained attempt at axiomatic analysis. We should anticipate interesting results of the sort that should contribute to the development of better heuristics in both theorem proving as it is actually done informally and in problem solving as it is now done in the quantitative and mathematically oriented empirical sciences.

4. CONCLUDING REMARK ON THE DISTINCTION BETWEEN HEURISTICS AND AXIOMATICS

It is possible to take a line that says that what I have urged in the preceding section blurs the distinction between the axiomatic analysis of informal proofs and derivations and the codification of heuristics for the activities of giving such informal proofs and derivations. It is possible to take this line but I think it is important to maintain the distinction. The axiomatic analysis of qualitative derivations of differential equations in a given domain can aim at a kind of informal rigor characteristic of contemporary mathematics, especially characteristic of the kinds of discussions of such matters in modern probability theory, for example. On the other hand, I see the development of heuristics as being more psychological and much more incomplete. The real problem is to develop a useful heuristics that is, on the one hand, not a collection of general banalities but, on the other hand, is not simply an axiomatic analysis of the process that is the focus of inquiry.

To make this distinction more vivid, let me consider an example in my own experience. I am responsible at Stanford for a computer-assisted instruction course in axiomatic set theory. This is an intermediate undergraduate course giving students their first introduction to the subject. Proofs are given in an informal style but the computer program the students address in giving their proofs must construct internally a formal representation. We also have for certain parts of the course, and we hope in the future to have in more parts, hints that are given to the student about constructing a proof of a given theorem. Our objective is to have contingent hints that are based on an analysis of the student's proof and that give him advice on how to complete it. In constructing a computer program able to give such contingent hints we have no intention of being able to provide an adequate analysis of every proof a student might give. Some of the theorems are rather hard and some of the partial proofs will be too deviant for the heuristic program to understand them. In contrast, the informal proof procedures are meant to be complete in the sense that a student knows that he has available machinery adequate to giving the proof. Moreover, we know from our own experience that a variety of proofs can be constructed for any of the theorems assigned by using the informal proof procedures available. I would expect this contrast to continue. It is why I think a deeper theory of heuristics than anything I have suggested should be to a large extent psychological in character.

Another way of putting the matter is that a virtue of the axiomatic method is that it brings an unusual and sometimes startling degree of explicitness to the analysis of a subject matter, and I do not think of heuristics as doing this. If a heuristic achieves a total degree of explicitness, it passes from being a heuristic to being an algorithm. The contrast I have in mind, put still another way, is that axiomatic analysis primarily deals with the analysis of a subject matter. Heuristics should deal with a process or activity. We are as incomplete in the formulation of heuristics as we are incomplete in the formulation of rules for learning or performing any finely tuned skill. Readers of Polya can increase their skill in problem solving just as readers of a good manual on tennis can improve their game. But in both cases the rules that are formulated are only hints at how the skill should in fact be exercised. From simply reading the statement of rules about serving or hitting a good backhand in tennis it would be impossible, in fact, to play the game well. Not only are the rules quite incomplete in statement, but a person must actually practice in a nonverbal and active way the skills themselves in order to acquire any competence. Exactly the same thing, it seems to me, is true of heuristics. We cannot hope to teach each other at a deep level how to discover new theorems or to solve new problems in any detailed way. We can only provide heuristics to point in certain directions that make us perform more efficiently and more effectively. The rules of heuristics are as incomplete, fragmentary, and insufficient as are manuals of any other skill, from tennis to glass blowing.

6

REPRESENTATION THEORY AND THE ANALYSIS OF STRUCTURE

A central topic in the philosophy of science is the analysis of the structure of scientific theories. Much of my own work has been concerned with this topic, but in a particular guise. The fundamental approach I have advocated for a good many years is the analysis of the structure of a theory in terms of the models of the theory. In a general way, the best insight into the structure of a complex theory is by seeking representation theorems for its models, for the syntactic structure of a complex theory ordinarily offers little insight into the nature of the theory. I develop that idea here in a general way, and expand upon things I have written earlier. I begin with some informal introductory remarks about the nature of representations. The first section is devoted to the central concept of isomorphism of models of a theory, the second section to the nature of representation theorems, with some elementary examples given, and the third section to the related question of invariance and meaningfulness of a representation.

A *representation* of something is an image, model, or reproduction of that thing. References to representations are familiar and frequent in ordinary discourse.[1] Some typical instances are these:

*Reprinted from *Philosophia Naturalis*, **25** (1988), 254–268.

[1] Other meanings of representation will not be analyzed here, even though a close affinity can be found for many of them, as in 'The representation of the union approached management yesterday'.

> Sleep is a certain image and representation of death.
>
> The Play is a representation of a world I once knew well.
>
> It is the very representation of heaven on earth.
>
> The representation of Achilles in the painting was marvelous.
>
> This is a representation of the triumphal arch erected by Augustus.
>
> An intuitive and visual representation of nuclear forces is not possible.

In some cases we can think of a representation as improving our understanding of the object represented. Many of us certainly understand the proportions of a building better—especially the layout of the interior—after examining its architectural drawings.

The formal or mathematical theory of representation has as its primary goal such an enrichment of the understanding, although there are other goals of representation of nearly as great importance—for instance, the use of numerical representations of measurement procedures to make computations more efficient. Representation in the formal sense to be developed here has also been closely associated with reduction. An admirable goal accepted on almost all sides is to reduce the unknown to the known. Controversies arise when claims about reduction are ideological rather than scientific in character. It is usually not appreciated how involved and technical the actual reduction of one part of science—even a near neighbor—is to another.

Philosophical claims about the reduction—and thus representation—of one kind of phenomena or set of ideas by another are as old as philosophy itself. Here is Epicurus' reduction of everything to simple bodies, i.e., atoms, and space in his letter to his follower Herodotus:

> Moreover, the universe is bodies and space: for that bodies exist, sense itself witnesses in the experience of all men, and in accordance with the evidence of sense we must of necessity judge of the imperceptible by reasoning, as I have already said. And if there were not that which we term void and place and intangible existence, bodies would have nowhere to exist and nothing through which to move, as they are seen to move. And besides these two nothing can even be thought of either by conception or on the analogy of things conceivable such as could be grasped as whole existences and not spoken of as the accidents or properties of such existences. Furthermore, among bodies some are compounds, and others those of which

> compounds are formed. And these latter are indivisible and unalterable.

This passage from Epicurus, written about 300 B.C., is nearly duplicated in several places in Lucretius' long poem *De Rerum Natura* written about 250 years later. The reduction of all phenomena to the motion of atoms in the void was a central theme of ancient atomism, and the speculative development of the ideas of significance for the scientific developments that occurred much later.

A claimed reduction much closer to the formal spirit promoted here and one of great importance in the history of ideas is Descartes' reduction of geometry to algebra. He puts the matter this way in the opening lines of his *La Geometrie* (1637, 1954, p. 2):

> Any problem in geometry can easily be reduced to such terms that a knowledge of the lengths of certain straight lines is sufficient for its construction. Just as arithmetic consists of only four or five operations, namely, addition, subtraction, multiplication, division and the extraction of roots, which may be considered a kind of division, so in geometry, to find required lines it is merely necessary to add or subtract other lines; or else, taking one line which I shall call unity in order to relate it as closely as possible to numbers, and which arbitrarily, and having given two other lines, to find a fourth line which shall be to one of the given lines as the other is to unity...

The difference between these two theses of reduction could hardly be greater in the degree to which they were carried out at the time of their conception. The ancient atomists could establish in a satisfactory scientific sense practically nothing about their reductive thesis. Descartes' detailed mathematical treatment constituted one of the most important conceptual breakthroughs of early modern mathematics. On the other hand, Descartes' attempted reduction of matter to nothing but extension in his *Principles of Philosophy* (1644) was in its way just as speculative as that of Epicurus or Lucretius.

I emphasize that these comparisons are not meant to encourage a reductionistic methodology that asserts we should only talk about reductions that can be fully carried out from a formal standpoint. Nothing could be further from the truth. As an unreconstructed pluralist, I am happy to assign a place of honor to speculation as well as results, especially in view of how difficult it is to establish specific results on reduction for any advanced parts of science. We just need to recognize speculation for what it is.

1. ISOMORPHISM OF MODELS

One of the most general and useful set-theoretical notions that may be applied to a theory is the concept of two models of a theory being isomorphic. Roughly speaking, two models of a theory are isomorphic when they exhibit the same structure from the standpoint of the basic concepts of the theory. The point of the formal definition of isomorphism for a particular theory is to make this notion of *same structure* precise. It is to be emphasized, however, that the definition of isomorphism of models of a theory is not dependent on the detailed nature of the theory, but is in fact sufficiently independent often to be termed "axiom free." The use of the phrase "axiom free" indicates that the definition of isomorphism depends only on the set-theoretical character of models of a theory. Thus two theories whose models have the same set-theoretical character, but whose substantive axioms are quite different, would use the same definition of isomorphism.

These ideas may be made more definite by giving the definition of isomorphism for algebras that are often groups. Here a structure $(A, \circ, e, ^{-1})$ is an *algebra* if A is a nonempty set, $\circ$ is a binary operation from $A \times A$ to A, e is an element of A, and $^{-1}$ is a unary operation from A to A.

DEFINITION 1. *An algebra* $\mathfrak{A} = (A, \circ, e, ^{-1})$ *is* isomorphic *to an algebra* $\mathfrak{A}' = (A', \circ', e', ^{-1'})$ *if and only if there is a function f such that*

(i) *the domain of f is A and the range of f is A',*

(ii) *f is a one-one function,*

(iii) *if x and y are in A, then* $f(x \circ y) = f(x) \circ' f(y)$,

(iv) *if x is in A, then* $f(x^{-1}) = f(x)^{-1'}$,

(v) $f(e) = e'$.

When we ask ourselves whether or not two distinct objects have the same structure, we obviously ask relative to some set of concepts under which the objects fall. It is an easy matter to show that the relation of isomorphism just defined is an equivalence relation among algebras, i.e., it is reflexive, symmetric, and transitive. As a rather interesting example, we might consider two distinct but isomorphic groups which have application in the theory of measurement. Let one group be the additive group of integers. In this case, the set A is the set of all integers, the operation $\circ$ is the operation of addition, the identity element e is 0, and the inverse operation $^{-1}$ is the negative operation. As the second group, isomorphic to the first, consider the multiplicative group of all integer powers of 2. In this case, the set A' is the set of all numbers that are equal to 2 to some integer power, the operation $\circ'$ is the operation of multiplication, the identity element is the integer 1, and the inverse operation is the standard reciprocal operation, i.e., the inverse of x is $1/x$. To establish the isomorphism of the two groups $\mathfrak{A} = (A, +, 0, -)$ and $\mathfrak{A}' = (A', \cdot, 1, ^{-1})$, we may use the function f such that for every integer n in the set A

$$f(n) = 2^n \ .$$

Then it is easy to check that the range of f is A', that f is one-one, and

$$\begin{aligned} f(m \ \circ \ n) \ &= f(m+n) = 2^{m+n} = 2^m \cdot 2^n = f(m) \cdot f(n) \\ &= f(m) \ \circ' \ f(n), \\ f(n^{-1}) \ &= f(-n) = 2^{-n} = \tfrac{1}{2^n} = f(n)^{-1'}, \end{aligned}$$

and

$$f(0) = 2^0 = 1.$$

It should be apparent that the same isomorphism between additive and multiplicative groups is possible if we let the set of objects of the additive group be the set of all real numbers, positive or negative, and the set of objects of the multiplicative group be the set of all positive real numbers. From the standpoint of the theory of measurement, this isomorphism is of interest primarily because it means that there is no mathematical basis for choosing between additive and multiplicative representations. Standard discussions of extensive quantities, for example, those concerning the measurement of mass or distance, often do not emphasize that a multiplicative representation is as acceptable and correct as an additive representation. Because measurements of mass or distance are never negative, it may be thought that the remarks about groups do not apply precisely, for the additive groups considered all have negative numbers as elements of the group. The answer is that in considering the actual measurements of mass or distance, we restrict ourselves to the semigroup of positive elements of the additive group in question. However, the details of this point are not relevant here. Concerning the earlier remark that isomorphism or sameness of structure is relative to a set of concepts, note that the integers and the multiplicative group of powers of two differ in many number-theoretical properties.

As another simple example of a theory axiomatized by defining a set-theoretical predicate, we may consider the ordinal theory of measurement. Models of this theory are customarily called weak orderings and we shall use this terminology in defining the appropriate predicate.

The set-theoretical structure of models of this theory is a nonempty set A and a binary relation R defined on this set. Let us call such a couple $\mathfrak{A} = (A, R)$ a *simple relation structure.* We then have the following.[2]

DEFINITION 2. *A simple relation structure* $\mathfrak{A} = (A, R)$ *is a* weak ordering *if and only if for every* x, y, *and* z *in* A

(i) if xRy and yRz then xRz,

(ii) xRy or yRx.

The definition of isomorphism of simple relation structures should be apparent, but for the sake of explicitness I give it anyway, and emphasize once again that the definition of isomorphism depends only on the set-theoretical structure of the simple relation structures and not on any of the substantive axioms imposed.

DEFINITION 3. *A simple relation structure* $\mathfrak{A} = (A, R)$ *is* isomorphic *to a simple relation structure* $\mathfrak{A}' = (A', R')$ *if and only if there is a function* f *such that*

(i) *the domain of* f *is* A *and the range of* f *is* A',

(ii) f *is a one-one function,*

(iii) *if* x *and* y *are in* A *then* xRy *if and only if* $f(x)R'f(y)$.

To illustrate this definition of isomorphism let us consider the question, "Are any two finite weak orderings with the same number of elements isomorphic?" Intuitively it seems clear that the answer should be negative, because in one of the weak orderings all the objects could stand in the relation R to each other and not so in the other. It will be interesting to ask what is the counterexample with the smallest domain we can construct to show that such an isomorphism does not exist in general. It is clear at once that two one-element sets will not do, because within isomorphism there is only one weak ordering with a single element, namely the ordering that makes that single element stand in the given relation R to itself. However, a counterexample can be found by adding one more element. In one of the weak orderings we can let R be the universal relation, i.e., $R = A \times A$, the Cartesian product of A with itself, and in the other, let R'

[2] Notice that we use R to represent the weak ordering rather than the qualitative relation $\succsim$ much used elsewhere in this volume. The reason for this choice here is so as not to prejudge the direction of the ordering, which is intuitively implied by $\succsim$.

be a "minimal" relation satisfying the axioms for a weak ordering. More formally, let

$$\begin{aligned}
A &= \{1,2\}\\
R &= \{(1,1),(2,2),(1,2),(2,1)\}\\
A' &= A\\
R' &= \{(1,1),(2,2),(1,2)\}.
\end{aligned}$$

Then it is easily checked that $\mathfrak{A} = (A, R)$ and $\mathfrak{A}' = (A', R')$ are both weak orderings with domains of cardinality two, but A cannot be isomorphic to A'. For suppose there were a function f establishing such an isomorphism. Then we would have

$$1\ R\ 2 \text{ if and only if } f(1)\ R'\ f(2)$$

and

$$2\ R\ 1 \text{ if and only if } f(2)\ R'\ f(1),$$

but we also have 1 R 2 and 2 R 1, whence

$$f(1)\ R'\ f(2) \text{ and } f(2)\ R'\ f(1), \tag{1}$$

but this is impossible, for if $f(1) = 1$, then $f(2) = 2$, and thus from (1) 2 R' 1, but we do not have 2 R' 1. On the other hand, as the only other possible one-one function, if $f(1) = 2$ then $f(2) = 1$, and again we must have from (1) 2 R' 1, contrary to the definition of R'.

2. REPRESENTATION THEOREMS

In attempting to characterize the nature of the models of a theory the notion of isomorphism enters in a central way. Perhaps the best and strongest characterization of the models of a theory is expressed in terms of a significant representation theorem. By a *representation theorem* for a theory the following is meant. A certain class of models of a theory distinguished for some intuitively clear conceptual reason is shown to exemplify within isomorphism every model of the theory. More precisely, let **M** be the set of all models of a theory, and let **B** be some distinguished subset of **M**. A representation theorem for **M** with respect to **B** would consist of the assertion that given any model M in **M** there exists a model in **B** isomorphic to M. In other words from the standpoint of the theory every possible variation of model is exemplified within the restricted set **B**. It should be apparent that a trivial representation theorem can always be proved by taking $\mathbf{B} = \mathbf{M}$. A representation theorem is just as interesting as the intuitive significance of the class **B** of models and no more so. An example of a simple and beautiful representation theorem is Cayley's theorem that every group is isomorphic to a group of transformations. One source of the concept of a group, as it arose in the nineteenth century, comes from consideration of the one-one functions which map a set onto itself. Such functions are usually called transformations. It is interesting and surprising that the elementary axioms for groups are sufficient to characterize transformations in this abstract sense, namely, in the sense that any model of the axioms, i.e., any group, can be shown to be isomorphic to a group of transformations. (For a discussion and proof of this theorem, see Suppes, (1957), Ch. 12.)

Certain cases of representation theorems are of special interest. When the set **B** can be taken to be a unit set, i.e., a set with exactly one element, then the theory is said to be categorical. Put another way, a theory is categorical when any two models are isomorphic. Thus, a categorical theory has within isomorphism really only one model. Examples of categorical theories are the elementary theory of numbers when a standard notion of set is used, and the elementary theory of real numbers with the same standard notion of set. It has sometimes been asserted that one of the main differences between nineteenth- and twentieth-century mathematics is that nineteenth-century mathematics was concerned with categorical mathematical theories while the latter deals with noncategorical theories. It is doubtful that this distinction can be made historically, but there is certainly a rather sharp conceptual difference between working with categorical and noncategorical theories. There is a clear sense in which noncategorical theories are more abstract.

From a psychological standpoint a good case can probably be made for the view that a theory is regarded as abstract when the class of models becomes so large that any simple image or picture of a typical model is not possible. The range of models is too diverse; the theory is very noncategorical. Another closely related sense of "abstract" is that certain intuitive and perhaps often complex properties of the original model of the theory have been dropped, as in the case of groups, and we are now prepared to talk about models which satisfy a theory even though they have a much simpler internal structure than the original intuitive model. This meaning of "abstract" is very close to the etymological one.

Homomorphism of models. In many cases within pure mathematics a representation theorem in terms of isomorphism of models turns out to be less interesting than a representation theorem in terms of the weaker notion of homomorphism. A good example of this sort within the philosophy of science is provided by theories of measurement, and the generalization from isomorphism to homomorphism can be illustrated in this context. When we consider general practices of measurement it is evident that in terms of the structural notion of isomorphism we would, roughly speaking, think of the isomorphism as being established between an empirical model of the theory of measurement and a numerical model. By an *empirical model* we mean a model in which the basic set is a set of empirical objects and by a *numerical model* one in which the basic set is a set of numbers. However, a slightly more detailed examination of the question indicates that difficulties about isomorphism quickly arise. In all too many cases of measurement, distinct physical objects are assigned the same number, and thus the one-one relationship required for isomorphism of models is destroyed. Fortunately, this weakening of the one-one requirement for isomorphism is the only respect in which we must change the general notion, in order to obtain an adequate account for theories of measurement of the relation between empirical and numerical models. The general notion of homomorphism is designed to accommodate exactly this situation. To obtain the formal definition of homomorphism for two algebras or two simple relation structures as previously defined, we need only drop the requirement that the function establishing the isomorphism be one-one. When this function is many-one but not one-one, we have a homomorphism that is not an isomorphism.[3]

These remarks may be made more concrete by considering the theory of weak orderings as a theory of measurement. It is easy to give a simple example of two weak orderings such that the first is homomorphic to the second, but not isomorphic to it. Let

$$\begin{aligned} A &= \{1,2\} \\ R &= \{(1,1),(2,2),(1,2),(2,1)\} \\ A' &= \{1\} \\ R' &= \{(1,1)\} \end{aligned}$$

and

$$f(1) = 1$$

[3] A weaker notion of homomorphism is generally used in algebra. The condition that, e.g., structures (A, R) and (A', R') be homomorphic with f being the mapping A onto A' is that if xRy then $f(x)R'f(y)$, rather than if and only if. However, in the theory of measurement and in other applications in the philosophy of science, the definition used here is more satisfactory.

$$f(2) = 1.$$

From these definitions it is at once obvious that the weak ordering $\mathfrak{A} = (A, R)$ is homomorphic under the function f to the weak ordering $\mathfrak{A}' = (A', R')$.

The point in showing the homomorphism is that we have

$$1\ R\ 2 \text{ if and only if } f(1)\ R'\ f(2),$$

as well as

$$2\ R\ 1 \text{ if and only if } f(2)\ R'\ f(1),$$

and both these equivalences hold just because

$$f(1) = f(2) = 1.$$

On the other hand, it is also clear simply on the basis of cardinality considerations that $\mathfrak{A}$ is not isomorphic to $\mathfrak{A}'$, because the set A has two elements and the set A' has one element. It is also evident that $\mathfrak{A}'$ is not homomorphic to $\mathfrak{A}$. This also follows from cardinality considerations, for there is no function whose domain is the set A' and whose range is the set A. As this example illustrates, the relation of homomorphism between models of a theory is not an equivalence relation; it is reflexive and transitive, but not symmetric.

By a *numerical* weak ordering I mean a weak ordering $\mathfrak{A} = (A, \leq)$ where A is a set of numbers. The selection of the numerical relation $\leq$ to represent the relation R in a weak ordering is arbitrary, in the sense that the numerical relation $\geq$ could just as well have been chosen. However, choice of one of the two relations $\leq$ or $\geq$ is the only intuitively sound possibility. The following theorem provides a homomorphic representation theorem for finite weak orderings, and thus makes the theory of finite weak orderings a theory of measurement.

THEOREM 1. *Every finite weak ordering is homomorphic to a numerical weak ordering.*

Proof. Let $\mathfrak{A} = (A, R)$ be a finite weak ordering.
Probably the simplest approach is first to form equivalence classes of objects in A, with respect to the obvious equivalence relation E defined in terms of R:

$$xEy \text{ if and only if } xRy \ \&\ yRx.$$

Thus, using the standard notation "$[x]$" for equivalence classes, i.e.,

$$[x] = \{y : y \in A \,\&\, xEy\},$$

we first order the equivalence classes according to R. Explicitly, we define

$$[x]R^*[y] \text{ if and only if } xRy.$$

It is straightforward to prove that R^* is reflexive, antisymmetrical, transitive, and connected in the set A/E of equivalence classes, or, in other words, that it is a simple ordering of A/E. Since A is a finite set, necessarily A/E is finite. Let $[x_1]$ be the first element of A/E under the ordering R^*, $[x_2]$ the second,..., and $[x_n]$ the last element. Consider now the numerical function g defined on A/E, defined as follows:

$$g([x_i]) = i \text{ for } i = 1, \dots, n.$$

Then the function g establishes an isomorphism between the ordering $\mathfrak{A}/E = (A/E, R^*)$ and the numerical ordering $\mathfrak{N} = (N, \leq)$, where N is the set of first n positive integers. (The details of this part of the proof are tedious but obvious.) We then define the numerical function f on A, for every y in A, by:

$$f(y) = i \text{ if and only if } y \in [x_i],$$

i.e., if y is in the i^{th} equivalence class under the ordering R^* . The function f establishes a homomorphism between $\mathfrak{A}$ and $\mathfrak{N}$, as desired.

Theorem 1 was restricted to finite weak orderings for good reason; it is false if this restriction is removed. The classic counterexample is the lexicographical ordering of the plane.

Let A be the set of all ordered pairs (x, y) of real numbers, and let the relation R be defined by the equivalence $(x_1, x_2)\ R\ (y_1, y_2)$ if and only if $x_1 < y_1$, or $x_1 = y_1$ and $x_2 \leq y_2$. Suppose that there exists a real-valued function f satisfying the equivalence:

$$f(x) \leq f(y) \text{ if and only if } xRy. \tag{1}$$

We fix x_2 and y_2 with $x_2 < y_2$ and define for each x_1:

$$f'(x_1) = f(x_1, x_2)$$

$$f''(x_1) = f(x_1, y_2).$$

In terms of these functions define the following function g from real numbers to intervals:

$$g(x_1) = [f'(x_1), f''(x_1)].$$

On the assumption that the ordering is lexicographic, g must be one-one since two distinct numbers are mapped into two disjoint intervals. For instance, if $x_1 > x_1'$ then $f'(x_1) = f(x_1, x_2) > f(x_1', y_2) = f''(x_1')$. But it is well known that there can be no one-one correspondence between the uncountable set of real numbers and the countable set of nondegenerate disjoint intervals. Thus no such function g can exist, and a fortiori there can be no function f satisfying (1) for the lexicographic ordering.

Embedding of models. We have seen that the notion of two models being homomorphic is a generalization of the notion of two models being isomorphic. A still more general and therefore weaker relation between models is that of one model being embedded in another. To prove an embedding theorem for a theory is to prove that there is an interesting class $\boldsymbol{M}$ of models such that every model of the theory is isomorphic, or at least homomorphic, to a submodel belonging to $\boldsymbol{M}$. The exact definition of submodel will vary slightly from one theory to another depending on the set-theoretical character of its models. For example, if $\mathfrak{A} = (A, \circ, e, ^{-1})$ is an algebra as defined above, then an algebra $\mathfrak{A}' = (A', \circ', e', ^{-1'})$ is a subalgebra of $\mathfrak{A}$ if A' is a subset of $A, \circ'$ is the operation $\circ$ restricted to A' (i.e., $\circ' = \circ \cap (A' \times A' \times A')$), $e' = e$, and $^{-1'}$ is the operation $^{-1}$ restricted to A'. In the case of simple relation structures the definition is still simpler. Let $\mathfrak{A} = (A, R)$ and $\mathfrak{A}' = (A', R')$ be two such structures. Then $\mathfrak{A}'$ is a submodel of $\mathfrak{A}$ if A' is a subset of A and R' is the relation R restricted to A', i.e., $R' = R \cap (A \times A)$.

Theorem 1 could have been formulated as an embedding theorem along the following lines. Let Re be the set of real numbers. Then it is apparent at once that $(Re, \leq)$ is a numerical weak ordering as defined earlier, and every finite weak ordering can be homomorphically embedded in $(Re, \leq)$, i.e., is homomorphic to a submodel of $(Re, \leq)$.

3. INVARIANCE AND MEANINGFULNESS

In connection with any measured property of an object, or set of objects, it may be asked how unique is the number assigned to measure the property. For example, the mass of a pebble may be measured in grams or pounds. The number assigned to measure mass is unique once a unit has been chosen. A more technical way of putting this is that the measurement of mass is unique up to a similarity transformation.[4]

The measurement of temperature in °C or °F has different characteristics. Here an origin as well as a unit is arbitrarily chosen: technically speaking, the measurement of temperature is unique up to a linear transformation.[5] Other formally different kinds of measurement are exemplified by (1) the measurement of probability, which is absolutely unique (i.e., unique up to the identity transformation), and (2) the ordinal measurement of such physical properties as hardness of minerals, or such psychological properties as intelligence and racial prejudice. Ordinal measurements are commonly said to be unique up to a monotone increasing transformation.[6]

Use of these different kinds of transformations is basic to the main idea of this section. An empirical hypothesis, or any statement in fact, which uses numerical quantities is empirically meaningful only if its truth value is invariant under the appropriate transformations of the numerical quantities involved. As an example, suppose a psychologist has an ordinal measure of I.Q., and he thinks that scores $S(a)$ on a certain new test T have ordinal significance in ranking the intellectual ability of people. Suppose further that he is able to obtain the ages $A(a)$ of his subjects. The question then is: Should he regard the following hypothesis as empirically meaningful?

HYPOTHESIS 1. *For any subjects a and b, if* $S(a)/A(a) < S(b)/A(b)$, *then* I.Q.$(a) <$ I.Q. (b).

From the standpoint of the invariance characterization of empirical meaning, the answer is negative. To see this, let *I.Q.* $(a) \geq$ *I.Q.* (b), *let* $A(a) = 7$, $A(b) = 12$, $S(a) = 3$, $S(b) = 7$. Make no transformations on the I.Q. data, and make no transformations on the age data. But let ϕ be a monotone-increasing transformation which carries 3 into 6 and 7 into itself. Then we have

$$\frac{3}{7} < \frac{7}{12},$$

[4] A real-valued function ϕ is a *similarity* transformation if there is a positive number α such that for every real number x

$$\phi(x) = \alpha x.$$

In transforming from pounds to grams, for instance, the multiplicative factor α is 453.6.

[5] A real-valued function ϕ is a *linear* transformation if there are numbers α and β with $\alpha > 0$ such that for every number x

$$\phi(x) = \alpha x + \beta.$$

In transforming from Centigrade to Fahrenheit degrees of temperature, for instance, $\alpha = 9/5$ and $\beta = 32$.

[6] A real-valued function ϕ is a *monotone increasing* transformation if, for any two numbers x and y, if $x < y$, then $\phi(x) < \phi(y)$. Such transformations are also called *order-preserving*.

but

$$\frac{6}{7} \geq \frac{7}{12},$$

and the truth value of Hypothesis 1 is not invariant under ϕ.

The empirically significant thing about the transformation characteristic of a quantity is that it expresses in precise form how unique is the structural isomorphism between the empirical operations used to obtain a given measurement and the corresponding arithmetical operations or relations. If, for example, the empirical operation is simply that of ordering a set of objects according to some characteristic, then the corresponding arithmetical relation is that of less than (or greater than), and any two functions which map the objects into numbers in a manner preserving the empirical ordering are adequate. More exactly, a function f is adequate if, and only if, for any two objects a and b in the set, a stands in the given empirical relation to b if and only if

$$f(a) < f(b)\ .$$[7]

It is then easy to show that, if f_1 and f_2 are adequate in this sense, then they are related by a monotone-increasing transformation. Only those arithmetical operations and relations which are invariant under monotone-increasing transformations have any empirical significance in this situation.

When we turn from the examination of numerical quantities to models of more complex theories, we obtain results of a similar character. For example, in examining classical mechanics we get a representation that is unique, when units of measurement are fixed, up to a Galilean transformation, that is, a transformation to some other inertial system. In the case of relativistic structures of particle mechanics, the uniqueness is up to Lorentz transformations.

To give an example of an elementary result, we can state the uniqueness theorem corresponding to the representation theorem (Theorem 1) for finite weak orders.

THEOREM 2. *Let* $\mathfrak{A} = (A, R)$ *be a finite weak order. Then any two numerical weak orderings to which it is homomorphic are related by a strictly increasing numerical function.*

Put in other language, the numerical representation of finite weak orders is unique up to an ordinal transformation. Invariance up to ordinal transformations is not a very strong property of a measurement, and it is for

[7] For simplicity we shall consider here only the arithmetical relation $<$. There is no other reason for excluding $>$.

this reason that Hypothesis 1 turned out not to be meaningful, because the hypothesis was not invariant under monotone transformations of the measurement data.

I have not mentioned as yet what is probably the most important, certainly the most important historical domain, in which invariance and meaningfulness were applied, namely, geometry. Here is a famous quotation from Felix Klein from his Erlangen address of 1872 (see Klein, 1893—I have made occasional minor changes in the quotation of the English translation).

> For geometric properties are, from their very idea, independent of the position occupied in space by the configuration in question, of its absolute magnitude, and finally of the sense in which its parts are arranged. The properties of a configuration remain therefore unchanged by any notions of space, by transformation into similar configurations, by transformation into symmetrical configurations with regard to a plane (reflection), as well as by any combination of these transformations. The totality of all these transformations we designate as the *principal group* of space-transformations: *geometric properties are not changed by the transformations of the principal group*. And, conversely, *geometric properties are characterized by their remaining invariant under the transformations of the principal group*. For, if we regard space for the moment as immovable, etc., as a rigid manifold, then every figure has an individual character; of all the properties possessed by it as an individual, only the properly geometric ones are preserved in the transformations of the principal group. (p. 218)

Thus, under Klein's view, which is now widely adopted, one can recognize a meaningful Euclidean relation between points just by testing whether or not the relation is invariant under the group of Euclidean motions. Correspondingly, one can tell whether a relation between points is topologically meaningful by determining whether the relation is invariant under any homomorphism.

Moving back to the general scheme of things, a representation theorem should ordinarily be accompanied by a matching invariance theorem stating the degree to which a representation of a structure is unique. In the mathematically simple and direct cases it is easy to identify the group as some well-known group of transformations. For more complicated structures, for example, structures that satisfy the axioms of a scientific theory, it may be necessary to introduce more complicated apparatus, but the objective is the same, to wit, to characterize meaningful concepts in terms of invariance.

One note to avoid confusion: it is when the concepts are given in terms of the representation, for example, a numerical representation in the case of measurement, or representation in terms of Cartesian coordinates in the case of geometry, that the test for invariance is needed. When purely qualitative relations are given which are defined in terms of the qualitative primitives of a theory, for example, those of Euclidean geometry, then it follows at once that the defined relations are invariant and therefore meaningful. On the other hand, the great importance of the representations and the reduction in computations and notation they achieve, as well as understanding of structure, make it imperative that we have a clear understanding of invariance and meaningfulness for representations which may be in appearance, rather far removed from the qualitative structures that constitute models of the theory.

In the case of physics, the primitive notions themselves of a theory are not necessarily invariant. For example, if we axiomatize mechanics in a given frame of reference, then the notion of position for a particle, for example, is not invariant but is subject to a transformation itself. A more complicated analysis of invariance and meaningfulness is then required in such cases. The general point is clear, however: the study of representation is incomplete without an accompanying study of invariance of representation.

PART II

CAUSALITY AND EXPLANATION

7

CAUSAL ANALYSIS OF HIDDEN VARIABLES

My contribution to this symposium is focused on the retreat from strong conditions of causality that have been forced upon us by quantum mechanics. My intent is to describe in more or less successive stages the retreat from the paradise of deterministic causation. This retreat has taken place through a thicket of quantum-mechanical details. It is my intention to describe the general principles involved but to refer to the literature for proofs and full technical elaboration, even of matters that are crucial to the conceptual development.

The history of the efforts to prove or disprove the possibility of hidden variables begins at least with von Neuman and includes important work by Kochen and Specker and others, but much of the recent analysis has centered around Bell's inequality and related results. In spite of its importance and significance I shall ignore this earlier history and begin with these recent discussions.

The experimental situation most referred to by Bell and others is a system in which we are generating two spin-1/2 particle initially in the singlet state. We measure the spin of each particle as it leaves the source, one going in one direction and the other in the opposite direction. The puzzles and paradoxes arise from the strong dependencies we find

*Reprinted from *PSA 1980*, Vol. 2 (ed. by P. Asquith and R. Giere), 1981, 563-571. East Lansing, MI: Philosophy of Science Association.

between the spins of the two particles that are spatially separated, but which originated from the source in the singlet state at the same time.

There are five assumptions about these systems that are essentially noncontroversial, which I shall state here but not formulate in an explicit mathematical fashion.

(i) *Axial symmetry.* For any direction of the measuring apparatus the expected spin is 0, where spin is measured by +1 and −1 for spin +1/2 and spin −1/2, respectively. Further, the expected product of the spin measurements is the same for different orientations of the measuring apparatuses, as long as the angle between the measuring apparatuses remains the same. By the angle between the measuring apparatuses we refer to the angle between the different orientations. Thus, one apparatus, the one on the left, for example, might be oriented up and the one on the right 90 degrees away, taken counterclockwise. Notice that the assumption of axial symmetry is just like a standard assumption about the isotropy of space.

(ii) *Opposite measurement for same orientation.* The correlation between the spin measurements is −1 if the two measuring apparatuses have the same orientation. This assumption is theoretically sound but in actual measurements the correlations obtained are not precisely −1. We shall weaken this assumption in some of the subsequent discussion.

(iii) *Independence of the hidden variable.* The expectation of any function of the hidden variable. which we shall in accordance with the literature call λ, is independent of the orientation of the measuring apparatus. It is generally agreed that the hidden variable which gives us a causal analysis of the motion of the two spin-1/2 particles should not itself be affected by the way in which we happen to orient the measuring apparatus.

(iv) *Locality.* The spin measurement obtained with one apparatus is independent of the orientation of the other measuring apparatus.

(v) *Quantum-mechanical correlations.* The quantum-mechanical covariance for the spin of the two particles, given the values +1 and −1 as stated above, is $-cos\ \theta$, where θ is the angle between the orientations of the two measuring apparatuses.

Using these assumptions, with some changes here and there, we now proceed to chronicle the retreat to ever weaker causal conditions.

1. DETERMINISTIC CAUSES

It is natural in the framework of classical physics to add to the assumptions just given above that, given the hidden variable λ and the orientation of the measuring apparatus, the result of the spin measurement should be determined uniquely. In other words, intuitively the hidden variable λ should be a deterministic cause.

It is shown in Bell (1964, 1966) and with particular clarity in Wigner (1970) that under these assumptions there can be no hidden variable, so that the search for deterministic causes is mistaken.

In an earlier paper, Zanotti and I (1976) weaken the deterministic assumption to conditional statistical independence, that is, to the assumption that the expectation of a product of the spin measurements, given λ and the orientation of the measuring apparatuses, is equal to the product of the expectations under the same conditions. Our argument is a straightforward probabilistic one. We first show that statistical independence, given λ, together with a correlation of -1, implies determinism. I mention this assumption of conditional statistical independence because I shall be returning to it throughout the paper.

Within the deterministic framework of classical physics, the negative results of Bell constitute in the minds of many people the most definitive refutation of the search to expand classical quantum mechanics into a more encompassing classical theory of deterministic causes.

2. DE FINETTI'S THEOREM

Given that the case for determinism is hopeless in the context of quantum mechanics, the first line of retreat is to look for causal hidden variables that render the correlated spin phenomenon conditionally independent. The general rubric here, of widespread application in modern statistics, is that a proper probabilistic causal analysis should render the phenomenological data statistically conditionally independent. This is precisely the role of a probabilistic common cause. There is a famous theorem of de Finetti's that looks as if it might have some application here because of the very general results about conditional independence.

Before stating the theorem, I need to say something about the principle of exchangeability. It is a principle of symmetry that has not been used in physics in any extended way. The principle was introduced by de Finetti to provide a natural alternative in the subjective theory of probability to objective theories of independence. The subjective aspects of the principle are of no importance here but only its strong form of symmetry. Here is a simple example to illustrate exchangeability. Suppose we flipped ten times a coin whose bias is unknown. Then the flips will not be independent because the outcomes of preceding flips will provide information about the probability of a head on the next flip. On the other hand, given the number of heads that occur in ten trials, the trials in which the heads occur are of no importance. In other words, we have permutational invariance in the sense that the probability of a sequence of ten outcomes with a fixed number of trials is the same regardless of exactly on which trials heads occur. Notice that exchangeability as a principle of symmetry radically reduces the number of probabilities that have to be determined. In the case of the ten flips of a coin, instead of considering 2^{10} sequences of possible outcomes we can reduce this to just 11, the probabilities of 0 to 10 heads. A little later I shall reintroduce the principle of exchangeability in the particular application of the spin experiments.

Given the principle of exchangeability, de Finetti's theorem may be stated in the following form: An infinite sequence of random variables is exchangeable if and only if there exists a random variable, which we may think of as causal, such that the random variables in the infinite sequence have identical conditional distributions and are conditionally independent given this causal random variable. (For those used to thinking of de Finetti's theorem in terms of mixtures of distributions, what the formulation I am referring to does is simply treat the weightings of the mixing as being identified with a causal random variable.) For infinite sequences of random variables, my interpretation of de Finetti's theorem is that exchangeability is equivalent to being able to find a causal mechanism that renders the random variables of the original phenomenon conditionally independent. As a simple example, take the case of flipping a coin that may have a bias. As has already been mentioned, we have exchangeability but not independence of the flips phenomenologically because, as the flips continue, their outcomes give us information for predicting future outcomes but if we know the causal random variable—in this case, the parameter of the bias—then we have conditional independence and the proper abstract causal account of the phenomenon. It is important to realize in these formulations that of course the causal mechanism identified is abstract and in general will in no sense be the fullest one possible. What is important about the theorem from the standpoint of causal concepts is that it shows the close relation between properly designed experiments for investigating causes of phenomena and the principle of exchangeability.

It has not always been recognized that de Finetti's theorem is a fundamental contribution to the theory of causality. An infinite sequence, of course, is approximated by large numbers of trials in actual experimentation. What the theorem shows is that if we have an experimental design in which exchangeability is satisfied phenomenologically, then we know on the basis of de Finetti's theorem alone that a common cause can be found that will render the phenomenological data conditionally independent—just as any good common cause should. Satisfaction of exchangeability is a nontrivial matter in experiments, but it is also important to recognize that the causal results implied by de Finetti's theorem are already a step away from a purely deterministic causal requirement. There is a good deal more to be said about de Finetti's theorem from the standpoint of the general theory of causality but I shall move on now to the case of two exchangeable events.

3. EXCHANGEABILITY IN THE SPIN EXPERIMENTS

The kind of symmetry expressed in the principle of exchangeability applies directly to the spin experiments that are a centerpiece of the literature surrounding Bell's results. This point is really uncontroversial and is an accepted part of the phenomenological data of the spin experiments. To be completely explicit it will be useful to express exchangeability in a formal way. Let $\mathbf{X}$ be the random variable for the measuring apparatus on the left—and on occasion we will call it apparatus I—and let $\mathbf{Y}$ be the random variable for expressing the measurement on the right with apparatus II. Exchangeability of $\mathbf{X}$ and $\mathbf{Y}$ may then be expressed as follows:

$$P(\mathbf{X} = 1, \mathbf{Y} = -1) = P(\mathbf{X} = -1, \mathbf{Y} = 1).$$

The symmetry of experimental design immediately satisfies this principle for the spin experiments.

Now if we extend de Finetti's theorem to this much weaker and simpler situation of two exchangeable random variables, then we would want a causal hidden variable $\boldsymbol{\lambda}$ such that $\mathbf{X}$ and $\mathbf{Y}$ are conditionally independent, given $\boldsymbol{\lambda}$, and, secondly, the conditional distributions of $\mathbf{X}$ and $\mathbf{Y}$ are identical, given $\boldsymbol{\lambda}$. This second requirement of identity of conditional distributions is a fundamental aspect of de Finetti's result and a standard classical demand in the theory of causality. It is, in its own way, a theoretical principle of symmetry as opposed to the phenomenological principle of exchangeability. Thus, for example, when we throw out a pair of dice that have the same bias, that is, have the same causal hidden variable $\boldsymbol{\lambda}$, we expect identity of conditional distributions, namely, the conditional probability of a face is the same for both. The principle of symmetry is an old and classical one—there is no basis for the conditional distributions to be different. I emphasize a point that is sometimes forgotten in these discussions, that the actual outcomes will be different most of the time, in particular 5/6 of the times in the case of fair dice, *even though* the conditional distributions are identical.

Now for a second point about two exchangeable random variables such as $\mathbf{X}$ and $\mathbf{Y}$. It is easy to show that, in general for just two random variables as opposed to de Finetti's infinite sequence, an underlying causal hidden variable need not necessarily exist.

Zanotti and I (1980) proved the following theorem giving necessary and sufficient conditions on the phenomenological data for two exchangeable hidden variables such as $\mathbf{X}$ and $\mathbf{Y}$ to have a causal hidden variable that will render them conditionally independent with identical conditional distributions. The condition is that their correlation be nonnegative. I restate the theorem in the following more formal fashion.

THEOREM 1. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two-valued random variables, for definiteness with possible values 1 and* -1, *and with positive variances, i.e.,* $\sigma(\mathbf{X}), \sigma(\mathbf{Y}) > 0$. *In addition, let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be exchangeable. Then a necessary and sufficient condition that there exist a hidden variable* $\boldsymbol{\lambda}$ *such that* $E(\mathbf{XY}|= \boldsymbol{\lambda} = \lambda) = E(\mathbf{X}|\boldsymbol{\lambda} = \lambda)E(\mathbf{Y}|\boldsymbol{\lambda} = \lambda)$ *and* $E(\mathbf{X}|\boldsymbol{\lambda} = \lambda) = E(\mathbf{Y}|\boldsymbol{\lambda} = \lambda)$ *for every value* λ *(except possibly on a set of measure zero) is that the correlation of* $\mathbf{X}$ *and* $\mathbf{Y}$ *be nonnegative.*

Some related results about finite sequences of exchangeable random variables are to be found in Diaconis (1977). It is, as one might expect, easy to show that in the case of an infinite exchangeable sequence of random variables all pairs of random variables must necessarily have nonnegative correlation, so the condition that is imposed here is not one that is really stronger than one that holds for the infinite sequence of de Finetti's theorem. It is just that this condition now needs to be made explicit for the weaker case of two random variables.

From what was said at the beginning about negative correlations in the case of the spin experiments, it is obvious that the necessary and sufficient condition for the existence of a causal hidden variable $\boldsymbol{\lambda}$ will not be satisfied. Thus, in its most natural form our retreat from deterministic to probabilistic common causes that yield identical conditional distributions is not successful. Notice how little of quantum mechanics has been used in the present result—only the existence of negative correlations, not as in the case of Bell's earlier papers the specific covariance or correlation result for quantum mechanics in terms of the cosine of the angle between the orientation of the two apparatuses. What this theorem shows is that strong causal intuitions cannot be satisfied, even at the probabilistic level, in quantum mechanics. Something has to give and it must be either the requirement of conditional independence or the requirement of identity of conditional distribution. The first is a principle of locality and the second a natural principle of symmetry. There is a more detailed discussion of these matters in Suppes and Zanotti (1980). Here I continue the line of retreat.

4. PROBLEM OF MORE THAN TWO EXCHANGEABLE VARIABLES

As an illustration of how complicated the general theory of causality is, I mention the fact that there are no pretty and simple conditions now known in terms of the phenomenological data of pairwise covariances or correlations that guarantee an underlying causal hidden variable for n exchangeable variables, when $n > 2$. In other words, the pairwise interactions between the variables can assume a complicated pattern and it is not clear what are the natural necessary and sufficient conditions on this pattern to guarantee the existence of an underlying common cause. In other words, the right generalization of Theorem 1 for $n > 2$ is not at all obvious.

5. BELL'S STOCHASTIC INEQUALITY

Bell (1971) derived a useful and important inequality that requires no deterministic assumption. Let $\boldsymbol{A}$ and $\boldsymbol{A}'$ be two random variables corresponding to two orientations of the left apparatus and random variables $\mathbf{B}$ and $\mathbf{B}'$ be two orientations of the apparatus on the right in the spin experiments. Let us assume now that there is a causal hidden variable that renders the random variables conditionally independent but, I emphasize, does not necessarily guarantee identity of conditional distributions. Bell shows that the requirement of statistical conditional independence implies the following inequality:

$$(1) \qquad -2 \leq E(\mathbf{AB}) - E(\mathbf{AB}') + E(\mathbf{A}'\mathbf{B}) + E(\mathbf{A}'\mathbf{B}') \leq 2.$$

It is then easy to select angle values for the difference in orientation on the left and the right for the four expectations shown in this inequality such that the inequality is violated by the quantum-mechanical result (v) given above.

In our continual retreat, what we have now done is drop the theoretical symmetry of identical conditional distributions, kept only the locality condition of conditional independence, and yet, as Bell's inequality shows in the formulation he gives, there cannot exist an underlying common cause because of violation of his inequality. The central point for the present exposition is that still further retreat is required.

6. CAUSAL HIDDEN VARIABLES WITHOUT CONDITIONAL INDEPENDENCE

The results of Bell's inequality and the earlier results on exchangeability suggest that the best we can do is look for causal hidden variables that do not guarantee conditional independence but something less strong. If we look at examples in medicine and the social sciences where the search for conditional correlations is standard and the focus is the search for a common cause that factors out phenomenological correlations, it is absolutely standard not to expect to get results as strong as conditional independence. These applications of causal ideas are of course in highly empirical nontheoretical situations. The analysis does not take place in an environment where a strong fundamental theory is available.

A proper attitude about quantum mechanics is perhaps that it suggests a similar kind of result but at a deep theoretical level for physics. The demand for conditional independence is too strong a causal demand.

Unfortunately, once we give up conditional independence, within the framework of classical physics there is no obvious weaker but still quite general condition to impose on a causal theory for quantum and other phenomena. On the other hand, if we introduce relativistic considerations there is a natural way of expressing locality, namely, that if the state of the system is given just prior to the occurrence of an event of interest, no other earlier information about the system can change the conditional probability of the occurrence of the event in question. The so-called independence of path assumption is standard in stochastic processes and is easy to formulate in a relativistic setting. It prohibits, of course, instantaneous action at a distance and depends upon assuming that the propagation of any action cannot be faster than that of the velocity of light. It would take us beyond the framework of classical quantum mechanics to enter into this principle and I only mean to suggest that it is a way of finding a new line of retreat, hopefully one on which we can stand and move no further to the rear.

Some of the foundational discussions of quantum mechanics, both by philosophers and physicists, often imply, at least implicitly, that causal analysis of quantum phenomena is not really possible. Such a general conclusion seems to me clearly mistaken. We cannot have a causal theory of quantum phenomena as rich in structural properties as are the theories of nineteenth-century classical physics. Even the weaker but still powerful concept of a common probabilistic cause will not be usable without some changes. But causal notions are implicit in all systematic quantum phenomena, and I am confident that we will ultimately have a satisfactory general analysis of causal concepts applicable to quantum phenomena. Of course, by “satisfactory” I do not mean that all classical requirements will be met but rather that we shall have a concept of cause that is as strong and as complete as is consistent with current well-supported theories of quantum phenomena. The process of this clarification will undoubtedly have ramifications all the way back to ordinary talk about causes, and ultimately we shall have a new way of thinking about causes.

8

SCIENTIFIC CAUSAL TALK

It is a pleasure to reply to Martin's comments on my theory of probabilistic causality, for he raises issues that occur in a rather natural way and that no doubt have been of concern to others (Martin, 1981). I have divided my reply into four major topics, which I have organized in a different order from that of their occurrence in Martin's comments. The topics are: the problem of a unified language of causality, the role of set theory in science, the language of events in science and ordinary talk, and problems of intensionality.

1. PROBLEM OF A UNIFIED LANGUAGE OF CAUSALITY

Martin is concerned that the probabilistic theory I have introduced does not adequately account for both scientific and ordinary occurrences of causal terms. In my monograph (Suppes, 1970) I claimed that a unified account could be given. Ten years later I am less optimistic about this and I think I would accept his criticism that I did not really accomplish this task, and I would now agree it is a mistake to try to have a unified language of any tightness and completeness. I have become increasingly persuaded of the plurality of science and of other realms of experience (Suppes, 1981). There are, of course, common elements to scientific and ordinary talk; there is not some sharp division of the kind Carnap wanted,

[*]Reprinted from *Theory and Decision*, **13** (1981), 363–380. Written as a reply to Martin (1981)

for example, in his two senses of probability. Yet there is a great deal of diversity and no real reason to think that these diverse uses will tend to converge in the future. I shall give some detailed examples later in terms of the language of random variables.

It is clear that of the two directions my analysis of causal language might go it is more directed toward scientific practice. I do want to reemphasize that I do not think there is a sharp division between scientific talk and ordinary talk. In the article on the plurality of science I in fact argue for there being a veritable Tower of Scientific Babel, with each subdiscipline in science having its unique concepts and language. This is especially true of advanced experimental work. There is a common core of ordinary talk that almost all of us understand who speak English as a first language or as a highly developed second language. This core does not contain very much scientific language, but among subsets of speakers and listeners there is a common core of causal and probabilistic talk that goes smoothly over into more exact scientific talk. I shall not here try to chart that transition, which I think could in fact be documented empirically.

2. SET THEORY IN SCIENCE

One of Martin's points is that my use of a set-theoretical framework is mistaken, for such an apparatus is not needed scientifically. He proposes instead to use various philosophical variants, such as an axiomatized Boolean algebra or a language of part–whole as exemplified by mereology. I think he is flatly and unequivocally wrong. The idiosyncratic languages he talks about are of interest in philosophy but for very special reasons, and they cut off philosophical discourse about causality from the mainstream of scientific talk. It is worth noting that none of the more complicated set-theoretical machinery I consider is duplicated in any way by Martin—for instance, the detailed and extensive learning-theory example which uses a probability space of countable sequences, or the entire apparatus of random variables which is the standard apparatus in modern probability theory and modern mathematical and applied statistics. The kind of language moves that Martin proposes would lead to further isolation of philosophical talk about these matters, an isolation that has already been too noticeable in the literature on confirmation and the foundations of induction.

I am quite willing to accept that with enough effort, all of the standard machinery characteristic of modern probability theory, not to speak of statistical theory, could be built up in one of Martin's idiosyncratic frameworks. But this would seem to me to be a terrible waste of time, and at the same time would isolate the developments from the large and interesting literature in science, mathematics, and statistics on these matters.

To drive this point home, I would like to consider one extended example to show why set-theoretical apparatus is natural even if not necessary. The purpose of this example is to introduce standard mathematical concepts that are needed for a causal analysis but that would not be readily available in any of the language frameworks suggested by Martin.

We may take as an example of suitable complexity the theory of linear learning models set forth in Estes and Suppes (1959a). We assume that on every trial the organism can make exactly one of r responses, $A_i, i = 1, \ldots, r$ and that after each response it receives one of $r + 1$ reinforcements, $E_j, j = 0, 1, \ldots, r$. A learning parameter θ , which is a real number such that $0 < \theta \leq 1$, describes the rate of learning in a manner to be made definite in a moment. A possible realization of the theory is an ordered triple $\mathcal{X} = \langle X, P, \theta \rangle$ of the following sort. X is the set of all sequences or ordered pairs $\langle i, j \rangle$ of natural numbers with $i = 1, \ldots, r$ and $j = 0, 1, \ldots, r$. P is a probability measure on the smallest σ-algebra $\mathcal{B}(X)$ of cylinder sets of X, and θ is a real number as already described. (Cylinder sets are those events definable by the outcome of a finite number of trials.) To define the models of the theory, we need a certain amount of notation. Let $A_{j,n}$ be the event of response j on trial n; $E_{k,n}$ the event of reinforcement k on trial n, and for x in X, let $[x_n]$ be the equivalence class of all sequences in X that are identical with x through trial n, and let $P_{xj,n} = P(Aj, n|[x]_{n-l})$. We may then characterize the theory by the following set-theoretical definition.

DEFINITION. *A triple* $\mathcal{X} = \langle X, P, \theta \rangle$ *is a* linear learning model *if and only if the following three axioms are satisfied for every n, every x in X with* $P([x]_n) > 0$ *and every j and k:*

1. *If* $x \in E_{k,n}$ *and* $j = k$ *and* $k \neq 0$ *then*
$$P_{xj,n+1} = (1 - \theta)p_{xj,n} + \theta;$$
2. *If* $x \in E_{k,n}$ *and* $j \neq k$ *and* $k \neq 0$ *then*
$$P_{xj,n+1} = (1 - \theta)p_{xj,n};$$
3. *If* $x \in E_{0,n}$ *then*
$$P_{xj,n+1} = p_{xj,n}.$$

The three axioms express assumptions concerning the effects of reinforcement and nonreinforcement. The first two say, in effect, that when a reinforcing event occurs, the response class corresponding to it increases in probability and all others decrease. A similar assumption is utilized in a number of stochastic and statistical models of learning. The third axiom expresses the assumption that response probabilities are unchanged on nonreinforced trials.

The critical point for the present discussion is the characterization of the probability measure P. It is easy to show that three conceptual ingredients enter into determining uniquely the probability of any event's happening, for example, any response or response sequence. The first ingredient is the initial probability of response at the beginning of the experiment before any reinforcements have been delivered; the second is the learning parameter θ that determines how fast change in behavior takes place under various reinforcement schedules; and the third is the schedule of reinforcements, which in general will be probabilistic in character and contingent upon previous reinforcements or responses. These three ingredients are the three causal factors, and theoretically the *only* causal factors determining the probability measure P, which fixes the probability of any event. The quantitative causal relations between events are all in turn determined by the measure P.

Of course, what I have just given is an informal analysis. In order to make it clear that the apparatus Martin refers to is far too elementary, I give the statement of the theorem and its complete proof in an appendix. The theorem can be regarded as a theorem about causality in learning theory. The rather lengthy and somewhat technical developments seem unavoidable in establishing precise results about the causal structure of models of the theory. I note among other things that the proof depends upon the well-known theorem of topology that a decreasing sequence of nonempty compact sets has a nonempty intersection. Secondly, the theorem does not hold for a finitely additive measure on all subsets of X, but only on the σ-algebra $\mathcal{B}(X)$ of cylinder sets, a somewhat delicate set-theoretical point.

3. LANGUAGE OF EVENTS

There is a considerable area of agreement between Martin and me concerning what he has to say about my discussion of events and his objections to my analysis being in certain directions too simplified. The language of events comes into play in much ordinary talk and in many parts of science. I would again take a pluralistic view that it is probably not possible to give a tightly unified account of these many different uses.

First I want to make a couple of technical remarks in response to some things that Martin says. He objects to my restriction to instantaneous events, and I certainly agree that, in general, this is not adequate. I certainly agree that this simplification restricts the applications of the formal concepts I introduced. I took it that it would be feasible but technically somewhat complicated to make the extension to noninstantaneous events.

One way to put what is somewhat surprising about Martin's objections to my use of standard event-language is that he simply does not consider the standard usage. It is as if someone were writing a treatise on the foundations of physics and assumed for that purpose classical mathematics. Someone who objected to classical mathematics might then raise objections to this use in physics of classical mathematics, but for most purposes such a move would be regarded as rather strange. It is part of the pluralism of approach I have already urged that when we are doing the foundations of causality, we should not try at the same time to reform the standard concepts of probability theory. Reforming or changing the standard concepts of probability is, for other purposes, a useful matter, but it is not even useful when it is idiosyncratic in the way that Martin's discussion is. The kind of discussion and framework he suggests in terms of mereology simply isolates all such discussion from the standard development of probability theory, as I have already argued. I have labored the point, but it seems to me to be worth laboring because adopting Martin's recommendations would isolate philosophical discussion of causality and a consequence of that isolation would be consideration only of the most elementary points about causality.

Martin objects, in particular, to my use of negation. Here I simply again followed standard usage. Complementation of an event is complementation with respect to the sample space or probability space. Such set-theoretical complementation is meant to correspond to the absence of occurrence of an event. My treatment here is standard and, as Jane Austin would say, unexceptionable. It is certainly possible to argue that in the translation of some ordinary talk this particular approach will not work. Certainly for the most general setting we might want to cite the fact that the complement of a set is not defined in Zermelo-Fraenkel set theory. Another point from another direction is that when the apparatus of random variables is used, as I claim most standard usage actually adopts for detailed statistical work, there is no longer a natural concept of negation in terms of random variables but only by reference once again to the sample space on which the random variable is defined.

It is certainly true that, in a certain sense, the notion of event as having definite physical properties is treated in a rather cavalier fashion in standard probability theory. I take it the reason for this is mainly the desire for flexibility and generality. When we are concerned with more specific philosophical questions or more specific physical questions, as for example in a discussion of how one should use the concept of event in relativistic physics, we may want to say a good deal more. I certainly want to admit that such extensions are proper but I also want to make clear that I think the direction Martin takes the discussion is mistaken.

For a general theory of causality with any pretension to be useful in a wide variety of sciences, it seems mistaken to tie down the concept of event by more detailed assumptions, as for example the kind that are easily suggested by physical theories. I do not see the concept of event as used in theories of space-time, for example, being of any real use and therefore of value in the formulation of causal concepts in economics or sociology. It is tempting to state my own general metaphysical views on the concept of physical events and to try to support my claim that the proper space for representing such events is atomless, but this does not seem the proper occasion.

4. INTENSIONALITY AND PROCEDURAL SEMANTICS

Martin quotes my own admission that the standard set-theoretical framework of probability concepts I adopt does not give a fully satisfactory treatment of intensional matters, especially for subjective theories of probability. Free substitution of terms that are held to be identical in the extensional sense leads to contradictions in the standard fashion.

On this point I agree with Martin and I agree with my earlier self. I remain, however, firm in the conviction that the handling of these intensional matters is not important in the framework of developing a causal theory for scientific purposes. A full-blown apparatus to handle these matters in completely explicit fashion will be another step toward the isolation I have already spoken of.

On the other hand, I think Martin is right in insisting that such intensional matters can be important in the sensitive analysis of causal concepts as they are referred to in ordinary talk. My own approach to such matters is to pursue, not for this reason alone but more generally in the interest of psychological realism, a move from set-theoretical to procedural semantics (Suppes, 1980, 1982). Unfortunately, it does not seem practical to go into these rather intricate matters in the short space of this reply.

I have covered what seems to me are the main points that Martin dealt with in some detail. He makes some passing mention of my references to different interpretations of probability but he does not develop this theme, and it therefore seems appropriate not to go into a discussion here. I will affirm, however, what I said in the original monograph. It seems to me there is a place for different uses of probability concepts ranging from that of a purely theoretical measure used as illustrated in several examples in the monograph in the formulation of theory. There is also a purely experimental use where one restricts oneself at most to a Bayesian prior if at all, and the data that carry the day from a probabilistic standpoint are the relative frequencies obtained in the experiment. There also remains the possibility of a generally subjective interpretation. I do not think it is necessary or perhaps even useful to try to draw a sharp line between these various uses. It is part of my pluralistic attitude to expect them. It is important to identify certain core properties that we expect any interpretation of probability to have.

I have enjoyed writing this reply to the substantive objections Martin makes to my ideas about causality. As is obvious, we disagree on many issues, but I do not expect to be able to offer precise arguments that will be regarded by him or by others as decisive. I do not think the subject of causality is like that. It has a glorious history and will have, no doubt, a robust pluralistic future. I hope only to help keep future efforts at analysis from being too much diverted from the mainstream of science to idiosyncratic philosophical bayous.

5. APPENDIX

In empirical applications of the learning theory described in the main text, the term $p_{xj,n}$ is to be interpreted as the probability of response A_j for a particular subject on trial n. In principle, the values of $p_{xj,n}$ can be predicted for all sequences and all n, given $p_{j,1}, r$ and θ (see Theorem 1 below). In practice, however, it is impracticable to evaluate trial by trial probabilities for individual subjects, so in experimental tests of the model we usually deal only with the average value of $p_{xj,n}$ over all sequences terminating on a given trial, i.e., with $p_{j,n}$. The latter can be predicted for all n, given the values of $p_{j,1}, r$ and θ, and sufficient information concerning probabilities or reinforcement and nonreinforcement (see Theorem 2 below).

We now turn to the two general theorems mentioned. The first theorem says that if $p_{j,1}, r$ and θ are given, then $p_{xj,n}$ is determined for all sequences x and all trials n. In formulating the theorem we make this idea precise by considering two models of the theory for which $p_{j,1}, r$ and θ are the same.

THEOREM I. *Let $\mathcal{X} = \langle X, P, \theta\rangle$ and $\mathcal{X}' = \langle X, P', \theta\rangle$ be two linear models for simple learning such that $p_{j,1} = p'_{j,1}$. Then if $P([x]_{n-l}) > 0$ and*

$P'([x]_{n-l}) > 0$, *we have:*

$$p_{xj,n} = p'_{xj,n}.$$

Proof. Suppose the theorem is false. Let n be the smallest integer such that (for some j and x)

$$p_{xj,n} \neq p'_{xj,n} \tag{1}$$

(By hypothesis of the theorem, $n > 1$.) Now if

$$P([x]_{n-1}) > 0 \tag{2}$$

and

$$P'([x]_{n-1}) > 0, \tag{3}$$

then by our hypothesis on n we have:

$$p_{xj,n-1} = p'_{xj,n-1}. \tag{4}$$

There are now three cases to consider: $x \in E_{j,n}$, $x \in E_{k,n}$ with $k \neq j$ and $k \neq 0$, and $x \in E_{0,n}$. Since the proof is similar for all three cases, each requiring application of the appropriate one of the three axioms, we consider only the first case:

$$x \in E_{j,n}. \tag{5}$$

From (2), (3), (5) and Axiom 1 we infer immediately:

$$p_{xj,n} = (1-\theta)p_{xj,n-1} + \theta \quad p_{xj,n} = (1-\theta)p'_{xj,n-1} + \theta. \tag{6}$$

From (4) and (6) we conclude:

$$p_{xj,n} = p'_{xj,n},$$

which contradicts (1) and establishes our supposition as false.

The second theorem establishes the fundamental result that given the initial probabilities of response of the subject, and the conditional probabilities of reinforcement, then a unique model of simple learning is determined. Moreover, no restrictions on these probabilities are required to establish the theorem. The significant intuitive content of this last assertion is that the experimenter may conditionalize the probabilities of reinforcement upon preceding events of the sample space in whatever manner he pleases.

Some preliminary definitions and lemmas are needed. The third definition introduces the notion of an *experimenter's partition* of X. The intuitive idea is that the conditional probabilities of reinforcing events on trial n depend on any partition of the equivalence classes $[x]_{n-1}$ and responses on the nth trial.[1]

DEFINITION 1. $\Xi(n) = \{\xi$: *there is an* x *in* X *and a* j *such that*

$$\xi = [x]_{n-1} \cap A_{j,n}\}.$$

$\Xi(n)$ is the finest experimenter's partition of X which we can use on the nth trial. It is immediately obvious that

LEMMA 1. *For every* n, $\Xi(n)$ *is a partition of* X.

We now use $\Xi(n)$ to define the general notion of an experimenter's partition $H(n)$, but for this definition we explicitly need the notion of one partition of a set being finer than another. (The definition is so phrased that any partition is finer than itself.)

DEFINITION 2. *If* $\mathcal{A}$ *and* $\mathcal{B}$ *are partitions of* X, *then* $\mathcal{A}$ *is finer than* $\mathcal{B}$ *if, and only if, for every set* A *in* $\mathcal{A}$ *there is a set* B *in* $\mathcal{B}$ *such that* $A \subseteq B$.

We than have:

DEFINITION 3. $H(n)$ *is an experimenter's partition of* X *(at trial* n*) if, and only if,* $H(n)$ *is a partition of* X *and* $\Xi(n)$ *is finer than* $H(n)$.

Finally, we need a lemma which provides a recursive equation for $P([x]_n)$ in terms of a given experimenter's partition on trial n. Notice that (iv) of the hypothesis of the lemma is a condition controlled by the experimenter, not by the subject.

LEMMA 2. *Let* $H(n)$ *be an experimenter's partition of* X. *Let*

(i) $\eta \in H(n)$,
(ii) $[x]_n \subseteq A_{j,n} \cap E_{k,n} \cap \eta$,
(iii) $P(A_{j,n} \cap [x]_{n-1}) > 0$,
(iv) $P(E_{k,n} | A_{j,n} \cap [x]_{n-1}) = P(E_{k,n} | \eta)$.

Then

$$P([x]_n) = P(E_{k,n}|\eta) p_{xj,n} P([x]_{n-1}).$$

Proof. By (ii) of the hypothesis

$$P([x]_n) = P(E_{k,n} \cap A_{j,n} \cap [x]_{n-1}),$$

whence,

$$P([x]_n) = P(E_{k,n}|A_{j,n} \cap [x]_{n-1}) P(A_{j,n}|[x]_{n-1}) P([x]_{n-1}).$$

Applying (iii) and (iv) to the first term on the right and the definition of $P_{xj,n}$ to the second, we obtain the desired result.

[1] A partition of a nonempty set X is a family of pairwise disjoint, nonempty subsets of X whose union is equal to X.

We are now prepared to state and prove the uniqueness theorem. Regarding the notation of the theorem it may be helpful to keep in mind that $q_{j,1}$ is the *a priori* probability of making response j on the first trial, and $\gamma_{\eta k,n}$ is the conditional probability of reinforcing event k on trial n given the event η of an experimenter's partition $H(n)$. It should be obvious why we use the notation $q_{j,1}$ rather than $p_{j,1}$ (and at the beginning of the proof $q_{xj,n}$ rather than $p_{xj,n}$); namely, the function p is *defined* in terms of the measure P whose unique existence we are establishing.

THEOREM 2. *Let X be an r-response space and let θ be a real number in the interval $(0,1]$, and let the numbers $q_{j,1}$ be such that*

$$q_{j,1} \geq 0, \qquad \sum_{j=1}^{r} q_{j,1} = 1.$$

For every n let $H(n)$ be an experimenter's partition of X, and let γ be a function defined for every n and k and every $\eta \in H(n)$ such that

$$\gamma_{\eta k,n} \geq 0, \qquad \sum_{k=0}^{r} \gamma_{\eta k,n} = 1.$$

Then there exists a unique probability measure P on $\mathcal{B}(X)$ such that

(i) *$\langle X, p, \theta\rangle$ is a linear model of simple learning,*

(ii) $q_{j,1} = p_{j,1}$,

(iii) $\gamma_{\eta k,n} = P(E_{k,n}|\eta)$.

(iv) *If $\eta \in H(n)$ and W is an $n-1$ cylinder set such that $W \subseteq \eta$ and $P(W) > 0$ then $P(E_{k,n}|W) = P(E_{k,n}|\eta$.)*

Proof. We first define recursively a function q intuitively corresponding to p, i.e., $q_{xj,n} = p_{xj,n}$.

(1) $q_{xj,1} = q_{j,1}$

(2) $q_{xj,n} = (1-\theta)q_{xj,n-1} + \theta\delta(j, \mathcal{E}(x, n-1)) + \theta q_{xj,n-1}\delta(0, \mathcal{E}(x, n-1))$,

where δ is the usual Kronecker delta function:

$$\delta(j,k) = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k, \end{cases}$$

and

(3) $$\mathcal{E}(x,n) = k \text{ if and only if } [x]_n \subseteq E_{k,n}.$$

(In effect, (2) combines all three axioms of the theory into one to provide this recursive definition.)

For subsequent use we prove by induction that

$$\sum_j q_{xj,n} = 1. \tag{4}$$

For $n = 1$, the proof follows at once from (1) and the hypothesis of the theorem that

$$\sum_j q_{xj,1} = 1.$$

Suppose now that

$$\sum_j q_{xj,n-1} = 1.$$

There are two cases to consider. If $x \in E_{k,n}$ for some $k \neq 0$ then from (2) and (3) we have at once:

$$\begin{aligned}\sum_j q_{xj,n} &= \sum_j (1-\theta)q_{xj,n-1} + \theta \\ &= (1-\theta)\sum_j q_{xj,n-1} + \theta \\ &= (1-\theta) + \theta \\ &= 1.\end{aligned}$$

If $x \in E_{0,n}$, then

$$\begin{aligned}\sum_j q_{xj,n} &= \sum_j [(1-\theta)q_{xj,n-1} + \theta q_{xj,n-1}] \\ &= \sum_j q_{xj,n-1} \\ &= 1.\end{aligned}$$

Following Lemma 2 we now recursively define $P([x]_n)$ in terms of q and the function γ introduced in the hypothesis of the theorem.

$$\left\{\begin{array}{l} P([x]_1) = q_{j,1}\gamma_{\eta_1}\mathcal{E}_{(x,1),1} \\ P([x]_n) = \gamma_\eta \mathcal{E}_{(x,n),n} q_{xj',n-1} P([x]_{n-1}), \end{array}\right. \tag{5}$$

where

$$\begin{array}{ll} [x]_1 \subseteq A_{j,1} & [x]_1 \subseteq \eta_1 \in H(1) \\ [x]_n \subseteq A_{j',n} & [x]_n \subseteq \eta \in H(n). \end{array}$$

We first need to show that the function P may be extended in a well-defined manner to any cylinder set C. To this end we prove by induction that if

$$C = \bigcup_{i=1}^{m_1} [x_i]_{n_1} = \bigcup_{i=1}^{m_2} [y_i]_{n_2}$$

then

$$\sum_{i=1}^{m_1} P([x_i]_{n_1}) = \sum_{i=1}^{m_2} P([y_i]_{n_2}). \tag{6}$$

When $n_1 = n_2$ the proof is trivial. Without loss of generality we may assume that $n_1 < n_2$; i.e., there is a positive integer t such that $n_1 + t = n_2$. We proceed by induction on t. But first we observe that the family of sets $[x_i]_{n_1}$ constitutes a partition of C, as does the family of sets $[y_i]_{n_1+t}$, and the latter is a refinement of the former. Whence for each set $[x_i]_{n_1}$ there is a subset I of the first m_2 positive integers such that

$$[x_i]_{n_1} = \bigcup_{h \in I} [y_h]_{n_1+t}. \tag{7}$$

And on the basis of (7) to establish (6) it is obviously sufficient to show that

$$P([x_i]_{n_1}) = \sum_{h \in I} P([y_h]_{n_1+t}).$$

Now if $t = 1$ then

$$\begin{aligned} [x_i]_{n_1} &= [x_i]_{n_1} \cap \bigcup_j A_{j,n_1+1} \cap \bigcup_k E_{k,n_1+1} \\ &= \bigcup_j \bigcup_k ([x_i]_{n_1} \cap A_{j,n_1+1} \cap E_{k,n+1}) \\ &= \bigcup_{h \in I} [y_h]_{n_1+1}. \end{aligned}$$

Since for $h \in I, [y_h]_{n_1} = [x_i]_{n_1}$, we infer from the above and (5) that

$$\begin{aligned} \sum_{h \in I} P([y_h]_{n_1+1}) &= \sum_j \sum_k \gamma_{\eta k, n_1+1} q_{xj,n_1} P([x_i]_{n_1}) \\ &= \sum_j q_{xj,n_1} P([x_i]_{n_1}) \ \text{ by hypothesis on } \gamma \\ &= P([x_i]_{n_1}) \ \text{ by (4).} \end{aligned}$$

Suppose now that (6) holds for t. Then there are sets I_1 and I_2 of positive integers such that

$$\begin{aligned}[x_i]_{n_1} &= \bigcup_{h\in I_1} [y_h]_{n_1+t} \cap \bigcup_j A_{j,n_1+t+1} \cap \bigcup_k E_{k,n_1+t+1}\\ &= \bigcup_{g\in I_2} [z_g]_{n_1+t+1}.\end{aligned}$$

Since for each $g \in I_2$ there is an h in I_1 such that

$$[z_g]_{n_1+t} = [y_h]_{n_1+t},$$

similarly to the case for $t = 1$ we infer that

$$\begin{aligned}\sum_{g\in I_2} P([z_g]_{n_1+t+1}) &= \sum_j \sum_k \sum_{h\in I_1} \gamma_{\eta k,n_1+t+1} q_{xj,n_1+t} P([y_h]_{n_1+t})\\ &= \sum_{h\in I_1} P([y_h]_{n_1+t})\\ &= P([x_i]_{n_1}),\end{aligned}$$

by our inductive hypothesis, which completes the proof of (6) and justifies the extension of P to any cylinder set: if

$$C = \bigcup_{i=1}^{m} [x_i]_n$$

then

$$P(C) = \sum_{i=1}^{m} P([x_i]_n).^2 \tag{8}$$

We now want to show that P is a probability measure on the algebra of cylinder sets of X. Since the functions q and γ are non-negative it follows at once from (5) and (8) that the nonnegativity probability axiom is satisfied, i.e., for every cylinder set C, $P(C) \geq 0$.

Now it is easy to select a subset Y of X such that

$$X = \bigcup_{x\in Y} [x]_1,$$

[2] In using the notation

$$C = \bigcup_{i=1}^{m} [x_i]_n$$

we always assume that sets $[x_i]_n$ are distinct (and consequently pairwise disjoint in this case); otherwise the extension of P would be incorrect.

whence by virtue of (5) and (8)

$$\begin{aligned} P(X) = \sum_{x \in Y} P([x]_1) &= \sum_j \sum_k q_{j,1}\gamma_{\eta k,1} \\ &= \sum_j q_{j,1} \sum_k \gamma_{\eta k,1} \\ &= 1 \cdot 1 \\ &= 1 \end{aligned}$$

which establishes that $P(X) = 1$.

To verify finite additivity of the measure P, let C_1 and C_2 be two cylinder sets such that $C_1 \cap C_2 = 0$. Without loss of generality we may assume they are both non-empty n-cylinder sets, and we may represent them each by

$$\begin{aligned} C_1 &= \bigcup_{i=1}^{m_1} [x_i]_n \\ C_2 &= \bigcup_{h=m_1+1}^{m_2} [x_h]_n, \end{aligned}$$

and by hypothesis, for each $i = 1, \ldots, m_1$ and $h = m_1 + 1, \ldots, m_2$

$$[x_i]_n \cap [x_h]_n = 0.$$

Whence

$$\begin{aligned} P(C_1 \cup C_2) &= P\left(\bigcup_{i=1}^{m_2} [x_i]_n\right) \\ &= \sum_{i=1}^{m_2} P([x_i]_n) \\ &= \sum_{i=1}^{m_1} P([x_i]_n) + \sum_{h=m_1+1}^{m_2} P([x_h]_n) \\ &= P(C_1) + P(C_2). \end{aligned}$$

Now for countable additivity. Let $\langle C_1, C_2, \ldots, C_n, \ldots \rangle$ be a decreasing sequence of cylinder sets, that is,

$$C_{n+1} \subseteq C_n \tag{9}$$

and

$$\bigcap_{n=1}^{\infty} C_n = 0. \tag{10}$$

Suppose now that

$$\lim_{n\to\infty} P(C_n) \neq 0. \tag{11}$$

(This limit must exist since the sequence is bounded and monotone decreasing. The monotonicity follows from (9) and the properties of P already established.)

In fact, let

$$\lim_{n\to\infty} P(C_n) = s > 0.$$

Hence for every n

$$P(C_n) \geq s,$$

and it follows at once that

$$C_n \neq 0. \tag{12}$$

We now use a topological argument to show that

$$\bigcap_{n=1}^{\infty} C_n \neq 0,$$

contrary to (10). The idea is simple; details will be omitted to avoid too serious a diversion. We know that X is the countably infinite product of a finite set. Hence, every cylinder set of X is compact in the product topology of the discrete topology on this finite set; in particular for every n, C_n is compact. Also by virtue of (12) every C_n is non-empty. But it is a well-known theorem of topology that a decreasing sequence of non-empty compact sets has a non-empty intersection, which contradicts (5). Thus our supposition (11) is false and the measure P is continuous from above at zero, which implies countable additivity.

Finally, the unique extension of P to the σ-algebra of cylinder sets follows from the standard theorem on this extension (see Kolmogorov, 1933, p. 17). The verification that the measure P defined by (5), (8) and the extension just mentioned has properties (ii)–(iv) of the theorem is straightforward and will be omitted.

9

EXPLAINING THE UNPREDICTABLE

It has been said—and I was among those saying it—that any theory of explanation worth its salt should be able to make good predictions. If good predictions could not be made, the explanation could hardly count as serious. This is one more attempt at unification I now see as misplaced. I want to examine some principled reasons why the thrust for predictability was mistaken. I begin with the familiar sort of example of explanation, the kind that occurs repeatedly in analyses of the past.

Hume's (1879) long and leisurely discussion of Charles I in his *History of England* provides a number of excellent examples. Here is one in which he is discussing Charles' decision to take action against the Scots in 1639.

> So great was Charles' aversion to violent and sanguinary measures, and so strong his affection to his native kingdom, that it is probable the contest in his breast would be nearly equal between these laudable passions and his attachment to the hierarchy. The latter affection, however, prevailed for the time, and made him hasten those military preparations which he had projected for subduing the refractory spirit of the Scottish nation. (*History of England*, Volume V, p. 107)

*Reprinted from *Erkenntnis*, **22** (1985), 187-195. The first version of this article was given at the Pacific Division meeting of the American Philosophical Association at a symposium on models of explanation, March 24, 1984. I am indebted to Jens Erik Fenstad for a number of helpful comments on the earlier draft.

Hume faces the standard difficulty of assessing attitudes and attachments when there is any sort of complex issue at stake. Charles' conflict between loyalty to Scotland and attachment to the religious hierarchy has the kind of psychological instability that makes prediction impossible. But from our perspective of looking back on the past, we are satisfied—at least many of us are—by the explanation that Charles' religious ties and commitments won out. By saying that we are satisfied I do not mean to suggest any ultimate sense of satisfaction.

The central feature of this example, the instability of Charles' conflicting feelings that are nearly equally matched, is the source of drama in many important historical events, in the tensions surrounding private choices of colleges to attend, careers to follow, and spouses to wed. The importance of such conflict and instability in our lives is mirrored in the importance they assume in the novels, plays, and movies that both express and define our ways of feeling and talking.

This instability and unpredictability of human affairs are in no sense restricted to conflicts of feeling. The vicissitudes of politics and war have been recorded and analyzed since Thucydides. In that long tradition, almost without exception there have been sound attempts at explanation but scarcely any attention given to what seemed to be the impossible task of predicting the outcomes. There is, in fact, a general view of the matter that is not correct in every detail but that expresses a major truth. Real conflicts occur in human affairs when the outcomes are uncertain, because the forces controlling them are unstable. One-sided battles that are known in advance by all concerned parties to be such are the exception rather than the rule. Napoleon thought he could conquer Russia, and at least some of his generals believed him. Hitler and at least some of his generals thought they could conquer the world. After the fact we can easily see how foolish they were.

One view of the American adversarial system of justice is that only conflicts in the law that have unpredictable outcomes should reach the stage of being tried in court. When the facts and the law are clear, early settlement should be reached, because the clarity about the facts and the law makes the situation stable and the outcome predictable. This rule does not always work but it probably covers a substantial majority of instances of legal conflict. Of course I do not want to overplay the argument for stability as the reason for settlement prior to trial. Just the expenses alone of a trial push the parties for settlement even when the facts and the law taken together do not provide a stable view of what the nature of a settlement should be. All the same the stability of the facts and the law is an important ingredient in many cases of conflict. The conflict goes nowhere because of the sound advice of a good attorney who convinces his client to control his anger and ignore his ruffled feathers. Major institutions of our society are organized to a large extent to deal with the instability generated by conflict. If the phenomena in question were predictable, much of the need for the institutions would be eliminated. There may still be Utopian social planners that dream of eliminating conflict and tension in some ideally structured future society, but most of us are prepared to accept conflict as part of the human condition and to work on ways to minimize it locally without hope of eliminating it.

The difficulties of predicting outcomes in the kind of human situations I have been describing are familiar. The complexity and subtlety of human affairs are often singled out as features that make a science of human behavior impossible, at least a predictive science in the way in which parts of physics and chemistry are predictive. Moreover, given the absence of powerful predictive methods, there are those who go on to say that the behavior in question is not explainable. I have already indicated my difference from this view.

I now want to move on to my main point. There is, I claim, no major conceptual difference between the problems of explaining the unpredictable in human affairs and in non-human affairs. There are, it is true, many remarkable successes of prediction in the physical sciences of which we are all aware, but these few successes of principled science making principled predictions are, in many ways, misleading.

Let me begin my point with a couple of simple examples. Suppose we balance a spoon on a knife edge. With a little steadiness of hand and patience, this is something that any of us can do. The spoon comes approximately to rest, perhaps still oscillating a little up and down. Our problem is to predict which way the spoon will fall, to the left or the right side of the knife blade, when there is a slight disturbance from a passing truck or some other source. In most such situations we are quite unable to make successful predictions. If we conduct this experiment a hundred times we will probably find it difficult to differentiate the sequence of outcomes from that of the outcomes of flipping a fair coin a similar number of times. Of course, in each of the particular cases we may be prepared to offer a sound schematic explanation of why the spoon fell to the left or to the right, but we have no serious powers of prediction.

Let me take as a second example one that I have now discussed on more than one occasion. The example originates with Deborah Rosen and was reported in my 1970 monograph on causality. A golfer makes a birdie by accidentally hitting a limb of a tree at just the right angle. The birdie is made by the ball's proceeding to go into the cup after hitting the limb of the tree. This kind of example raises difficulties for probabilistic theories of causality of the kind I have advocated. I do not want to go into the difficulties at this point but rather to use this simple physical example as a clear instance of having a good sense of explanation of the phenomenon, but not having any powers of predicting it. A qualitative explanation is that the exact angle at which the ball hit the limb of the tree deflected it into the cup. We cover the difficulties of giving a quantitative explanation by the usual qualitative method of talking about the ball's hitting the limb at "just the right angle." Of course, this is elliptical for "hitting the ball at just the right angle, just the right velocity, and just the right spin." The point is that we do not feel there is any mystery about the ball's hitting the limb and then going into the cup. It is an event that we certainly did not anticipate and could not have anticipated, but after it has occurred we feel as comfortable as can be with our understanding of the event.

There is a principled way of describing our inability to predict the trajectory of the golf ball. The trajectory observed with the end result of the ball's going into the cup is a trajectory followed in an unstable environment. We cannot determine the values of parameters sufficiently precisely to predict the golf ball will hit the limb of the tree at just the right angle for bouncing into the cup because the right conditions of stability do not obtain. To put it in a familiar way, very small errors in the measurement of the initial conditions lead to significant variations in the trajectory—here *significant* means going or not going into the cup. Correspondingly, when intentions are pure and simple we can expect human behavior to be stable and predictable, but as soon as major conflicts arise, e.g., of the sort confronting Charles I, the knowledge of intentions is in and of itself of little predictive help, though possibly of great explanatory help after the fact. To put it in a summary way, Charles I facing the Scots and our golf ball share a common important feature of instability.

Here is another simple example of a physical system of the sort much studied under what is currently called *chaos* in classical dynamics. We have a simple discrete deterministic system consisting of a ball being rotated around the circumference of a fixed circle, each move "doubling" the last. The only uncertainty is that we do not know the initial position with complete precision. There is a small uncertainty not equal to zero in our knowledge of the starting position of the ball. Then, although the motion of the system is deterministic, with each iteration around the circumference of the ball the initial uncertainty expands. More particularly, it is easy to show that after n iterations the uncertainty will be 2^n times the initial uncertainty. So it is obvious that after a sufficient number of iterated moves the initial uncertainty expands to fill the entire circumference of the circle, and the location of the ball on the circle becomes completely unpredictable.[1] Though the system just described is slightly artificial, it is enormously simple, with very limited degrees of freedom. It is an excellent example of an unstable dynamical system—the instability coming from the fact that a small uncertainty in the initial conditions produces an arbitrarily large uncertainty in subsequent location.

The general principle that I am stressing in these analyses that stretch from the mental conflicts of Charles I to simple rotating balls is the presence of instability as the central feature that makes prediction impossible. Fortunately, after the events have occurred we can often give a reasonable explanation.

I do not want to suggest that the absence of stability as such is the only cause for failure of prediction. We can take the view that there is an absence of determinism itself as in probabilistic quantum phenomena and in other domains as well. It will suffice here to consider some simple quantum examples. Perhaps the best is that of radioactive decay. We cannot predict when a particle will decay. We observe the uneven intervals between the clicking of a Geiger counter. After the events of decay have occurred we offer an "explanation," namely, we have a probabilistic law of decay. There is no hope of making an exact prediction but we feel satisfied with the explanation. Why are the intervals irregular? They are irregular because the phenomena are governed in a fundamental sense by a probabilistic decay law. Don't ask for a better explanation—none is possible.

I do not mean to suggest that instability and randomness are the only causes of not being able to make predictions. I do suggest that they provide principled explanations of why many phenomena are not predictable and yet in one sense are explainable.

The point I want to emphasize is that instability is as present in purely physical systems as it is in those we think of as characteristically human. Our ability to explain but not predict human behavior is in the same gen-

[1] A more technical description of such a simple deterministic description goes like this. Instead of moving around a circle, we consider a first-order difference equation, which is a mapping of the unit interval into itself:

$$x_{n+1} = 2x_n \pmod 1,$$

where mod 1 means taking away the integer part so that x_{n+1}, lies in the unit interval. So if $x_1 = 2/3, x_2 = 1/3, x_3 = 2/3, x_4 = 1/3$, etc., and if $x'_1 = 2/3+$ e, $x'_2 = 1/3+$ 2e, $x'_3 = 2/3 +$ 4e, and in general

$$x'_{n+1} = \begin{cases} 1/3 + 2^n e & \pmod 1 \text{ for } n \text{ even} \\ 2/3 + 2^n e & \pmod 1 \text{ for } n \text{ odd} \end{cases}$$

the instability of this simple system is evident, for the initial difference in x_1 and x'_1, no matter how small, grows exponentially.

eral category as our ability to explain but not predict many physical phenomena. The underlying reasons for the inability to predict are the same. The concept of instability which accounts for many of these failures is one of the most neglected concepts in philosophy. We philosophers have as a matter of practice put too much emphasis on the contrast between deterministic and probabilistic phenomena. We have not emphasized enough the salient differences between stable and unstable phenomena. One can argue that the main sources of probabilistic or random behavior lie in instability. We might even want to hold the speculative thesis that the random behavior of quantum systems will itself in turn be explained by unstable behavior of classical dynamical systems. But whether this will take place or not, much ordinary phenomena of randomness in the macroscopic world can best be accounted for in terms of instability. This is true of the behavior of roulette wheels as much as it is of the turbulence of air or the splash of a baby's bath.

A disturbing example of instability is to be found in the theory of population growth. A reasonable hypothesis is that the rate of growth is proportional to the current size of the population. The exponential solution of this equation is unstable. This means that slight errors either in the initial population count or in the constant of proportionality for the breeding rate can cause large errors in prediction, quite apart from any other influences that might disturb the correctness of the equation.[2]

It is worth saying once more in somewhat more abstract terms the central meaning of stability in the theory of dynamical systems. What I paraphrase here is the classical Lyapunov condition for a stable solution.[3] The idea is straightforward and already stated once intuitively. A solution is Lyapunov-stable if two different trajectories keep arbitrarily close together as they arise from different initial conditions provided the initial conditions are sufficiently close. So the intuitive idea of stability is that

[2] The differential equation expressing that growth of population $\frac{dx}{dt}$ is proportional to present population is

$$\frac{dx}{dt} = ax,$$

and the solution is

$$x = be^{at},$$

which is Lyapunov unstable, as defined in Note 3.

[3] The classical Lyapunov condition for a system of ordinary differential equations is the following, which formalizes the intuitive description in the text. Let a system of differential equations

$$(1) \qquad \frac{dx_i}{dt} = f_i(x_1, \dots, x_n, t), \qquad i = 1, \dots, n$$

a trajectory can be known with any desired precision, given sufficiently small errors of measurement in determining the initial conditions. This is exactly what is not characteristic of instability. Very fine variations in the initial conditions of a roulette wheel, not to speak of variations in its motion produce very large differences in outcome. Namely, in almost identical conditions we have on one occasion a red and on another occasion a black outcome. Now I am not suggesting that this exact idea of the stability of a dynamical system can be applied to our analysis of the behavior of Charles I deciding what to do with the Scots. There is, however, an underlying and robust notion of stability that reflects the instability in his behavior in a faithful way, just as much as it reflects the instability of a roulette wheel and its resulting random behavior.

The general qualitative concept of stability is this. A process is stable if it is not disturbed by causes of small magnitude. Thus, a chair is stable if it cannot be easily pushed over. A political system is stable if it can withstand reasonably substantial shocks. A person is stable if he is not continually changing his views. More specifically, a person's belief in a given proposition is stable if it can only be changed by very substantial new evidence. We often say something similar about feelings. One of the features of a stable personality is constancy of feeling. The Lyapunov formal definition of stability can be put under this qualitative tent.

Some of the best and most sophisticated predictive science is about well-defined stable systems, but here I am interested in the opposite story. When a system is unstable we can predict its behavior very poorly. Yet in many instances we can still have satisfactory explanations of behavior. There are at least three kinds of explanation that may qualify as satisfactory analyses of unpredictable behavior. The first and most satisfying arises from having what is supported by prior evidence as a highly accurate quantitative and deterministic theory of the phenomena in question. Classical physics has constituted the most important collection of such theories. In the golf-ball example discussed earlier we feel completely confident that no new fundamental physical principles are needed to give an account of the ball's surprising trajectory. It was and will remain hopeless to accurately predict such trajectories with their salient but unexpected

be given. A solution $y_i(t), i = 1, \ldots, n$ of (1) with initial conditions $y_i(t_0)$ is a *Lyapunov stable solution* if for any real number $\epsilon > 0$ there is a real number $\delta > 0$ such that for each solution $x_i(t), i = 1, \ldots, n$, if

$$|x_i(t_0) - y_i(t_0)| < \delta, \quad i = 1, \ldots, n$$

then

$$|x_i(t) - y_i(t)| < \epsilon, \quad i = 1, \ldots, n$$

for all $t \geq t_0$.

qualitative properties. Our serenity, however, is principled. Classical mechanical systems that are unstable have unpredictable behavior but the physical principles that apply to them are just those that are highly successful in predicting the behavior of stable systems. The explanatory extrapolation from stable to unstable systems seems conceptually highly justified by prior extensive experience. Moreover, in cases of importance, we can often estimate relevant parameters after the fact. Such estimates increase our confidence in our explanatory powers. Ex post facto stress analyses of structural failures in airplanes, bridges, and buildings are good instances of what I have in mind.

Application of fundamental theory or quantitative estimate of parameters seems out of the question in the second kind of explanation of unpredictable behavior I consider. Here I have in mind familiar common sense psychological explanations of unpredictable behavior, exemplified in the passage from Hume about Charles I. Consider, for instance, a standard analysis of an election that was said to be "too close to call." A variety of techniques are applied after the fact to explain the result: the bad weather affected Democrats more than Republicans, the last-minute interview of one candidate went badly, the rise in interest rates the past two weeks hurt the Republican candidate, and so on and so on. Simple psychological hypotheses relate any one of these explanatory conditions to the behavior of voters. Most of us have faith in at least some of these explanations, but we have no illusion that they are derived from a fundamental theory of political behavior. We also recognize the instability of the outcomes and the consequent difficulty of prediction.

The third kind of explanation of unpredictable behavior does not apparently depend on instability but on randomness. As has already been noted, the random behavior of classical mechanical systems, roulette wheels, for example, can be attributed to instability, but this is not the case for quantum phenomena. In either case, however, the important point is that explanation cannot go behind some basic probabilistic law that assigns a probability distribution to the phenomena in question. We explain the irregular pattern of radioactive decay or other data by the probability law thought to govern the phenomena. Individual events, no matter how controlled the environment, cannot be predicted with accuracy. Yet at a certain level we feel we have explained the phenomena.

Chaos, the original confusion in which all the elements were mixed together, was personified by the Greeks as the most ancient of the gods. Now in the twentieth century, chaos has returned in force to attack that citadel of order and harmony, classical mechanics. We have come to recognize how rare and special are those physical systems whose behavior can be predicted in detail. The naivete and hopes of earlier years will not return. For many phenomena in many domains there are principled reasons to believe that we shall never be able to move from good explanations to good predictions.

10

CONFLICTING INTUITIONS ABOUT CAUSALITY

In this article I examine five kinds of conflicting intuitions about the nature of causality. The viewpoint is that of a probabilistic theory of causality, which I think is the right general framework for examining causal questions. It is not the purpose of this article to defend the general thesis in any depth but many of the particular points I make are meant to offer new lines of defense of such a probabilistic theory. To provide a conceptual framework for the analysis, I review briefly the more systematic aspects of the sort of probabilistic theory of causality I advocate. I first define the three notions of prima facie cause, spurious cause, and genuine cause. The technical details are worked out in an earlier monograph (Suppes, 1970) and are not repeated.

DEFINITION 1. *An event B is a* prima facie cause *of an event A if and only if (i) B occurs earlier than A, and (ii) the conditional probability of A occurring when B occurs is greater than the unconditional probability of A occurring.*

Here is a simple example of the application of Definition 1 to the study of the efficacy of inoculation against cholera (Greenwood & Yule 1915, cited in Kendall & Stuart 1961). I also discussed this example in

*Reprinted from, *Midwest Studies in Philosophy* **IX** (ed. by P. A. French, T. E. Uehling, Jr. and H. K. Wettstein), 1984, pp. 150-168.

This article overlaps with Chapter 3 of my *Probabilistic Metaphysics,* (1984).

my 1970 monograph. The data from the 818 cases studied are given in the accompanying tabulation.

	Not attacked	*Attacked*	*Totals*
Inoculated	276	3	279
Not inoculated	473	66	539
Totals	749	69	818

The data clearly show the prima facie efficacy of inoculation, for the mean probability of not being attacked is 749/818 = 0.912, whereas the conditional probability of not being attacked, given that an individual was inoculated, is 276/279 = 0.989. Here A is the event of not being attacked by cholera and B the event of being inoculated.

In many areas of active scientific investigation the probabilistic data are not so clear-cut, although they may be scientifically and statistically significant. I have selected one example concerning vitamin A intake and lung cancer to illustrate the point. The results are taken from Bjelke (1975). The sample of Norwegian males 45-75 years of age was drawn from the general population of Norway but included a special roster of men who had siblings that had migrated to the United States. In 1964, the sample reported their cigarette smoking habits. More than 90 percent of those surviving in 1967 completed a dietary questionnaire sufficiently detailed to permit an estimate of vitamin A intake. On January 1, 1968, of the original sample, 8,278 were alive. Their records were computer-matched against the records of the Cancer Registry of Norway as of March 1, 1973.

The sample was classified into two groups according to an index of vitamin A intake as inferred from the dietary questionnaire, with 2,642 classified as having low intake and 5,636 as not low—I am ignoring in this recapitulation many details about this index. There were for the sample, as of March 1, 1973, 19 proven cases of carcinomas other than adenocarcinomas, which we ignore for reasons too detailed to go into here. Of the 19 proven cases, 14, i.e., 74 percent occurred among the 32 percent of the sample—the 2,642, who had a low intake of vitamin A. Only 5 cases, i.e., 26 percent, occurred among the 68 percent of the sample who had a high intake of vitamin A. Let C be the event of having a lung carcinoma and let L be low intake of vitamin A. Then for the sample in question

$$P(C) = .0023 < P(C|L) = .0053.$$

Using Definition 1 we infer that low intake of vitamin A is a prima facie cause of lung cancer. The probabilities in question are small but the results suggest further scientific investigation of the proposition that high intake of vitamin A may help prevent lung cancer.

It is now widely accepted that cigarette smoking causes lung cancer, but as the present data show, the incidence of lung cancer in the general population is so small that it is a primary medical puzzle to explain why so few smokers do get lung cancer. This study is meant to be a contribution to solving this puzzle.

An important feature of this study is that the results are fragile enough to warrant much further investigation before any practical conclusion is drawn—such as the admonition to heavy smokers to eat lots of carrots. In my view, perhaps a majority of scientific studies of causal connections have a similar tentative character. It is mainly science far from the frontiers, much worked over and highly selected, that has clear and decisive results.

A common argument of those who oppose a probabilistic analysis of causality is to claim that it is not possible to distinguish genuine prima facie causes from spurious ones. This view is mistaken. Because in my sense spuriousness and genuineness are opposites, it will be sufficient to define spurious causes, and then to characterize *genuine* causes as prima facie causes that are not spurious.

For the definition of spurious causes, I introduce the concept of a partition at a given time of the possible space of events. A partition is just a collection of incompatible and exhaustive events. In the case where we have an explicit sample space, it is a collection of pairwise disjoint, nonempty sets whose union is the whole space. The intuitive idea is that a prima facie cause is spurious if there exists an earlier partition of events such that no matter which event of the partition occurs, the joint occurrence of B and the element of the partition yields the same conditional probability for the event A as does the occurrence of the element of the partition alone. To repeat this idea in slightly different language, we have:

DEFINITION 2. *An event B is a spurious cause of A if and only if B is a prima facie cause of A, and there is a partition of events earlier than B such that the conditional probability of A, given B and any element of the partition, is the same as the conditional probability of A, given just the element of the partition.*

The history of human folly is replete with belief in spurious causes. One of the most enduring is the belief in astrology. The better ancient defenses of astrology begin on sound empirical grounds, but they quickly wander into extrapolations that are unwarranted and that would provide upon deeper investigation excellent examples of spurious causes. Ptolemy's treatise on astrology, *Tetrabiblos,* begins with a sensible discussion of how the seasons, the weather, and the tides are influenced by the motions of the sun and the moon. But he then moves rapidly to the examination of what may be determined about the temperament and fortunes of a given individual. He proceeds to give genuinely fantastic explanations of the cultural characteristics of entire nations on the basis of their relation to the stars. Consider, for example, this passage:

> Of these same countries Britain, (Transalpine) Gaul, Germany, and Bastarnia are in closer familiarity with Aries and Mars. Therefore for the most part their inhabitants are fiercer, more headstrong, and bestial. But Italy, Apulia, (Cisalpine) Gaul, and Sicily have their familiarity with Leo and the sun; wherefore these peoples are more masterful, benevolent, and co-operative (63, Loeb edition).

Ptolemy is not an isolated example. It is worth remembering that Kepler was court astrologer in Prague, and Newton wrote more about theology than physics. In historical perspective, their fantasies about spurious causes are easy enough to perceive. It is a different matter when we ask ourselves about future attitudes toward such beliefs current in our own time.

The concept of causality has so many different kinds of applications and is at the same time such a universal part of the apparatus we use to analyze both scientific and ordinary experience that it is not surprising to have a variety of conflicting intuitions about its nature. I examine five examples of such conflict, but the list is in no sense inclusive. It would be easy to generate another dozen just from the literature of the last ten years.

1. SIMPSON'S PARADOX

Simpson (1951) showed that probability relationships of the kind exemplified by Definition 1 for prima facie causes can be reversed when a finer analysis of the data is considered. From the standpoint of the framework of this article, this is just a procedure for showing that a prima facie cause is a spurious cause, at least in the cases where the time ordering follows the definitions given. In Simpson's discussion of these matters and in the related literature, there has not been an explicit attention to temporal order, and I shall ignore it in my comments on the 'paradox'. There is an intuitively clear and much discussed example of sex bias in graduate admissions at Berkeley (Bickel, Hammel, & O'Connell, 1975). When data from the university as a whole were considered, there seemed to be good evidence that being male was a prima facie cause for being admitted to graduate school. In other words, there was a positive bias toward the admission of males and a negative bias toward the admission of females. On the other hand, when the data were examined department by department it tuned out that a majority of the departments did not show such a bias and in fact had a very weak bias toward female admission. The conflict in the data arose from the large number of female applications to departments that had a large number of rejections independent of the sex of the applicant. As is clear from this example, there is no genuine paradox in the problem posed by Simpson. There is nothing inconsistent, or in fact even close to inconsistent, in the results described, which are characteristic of the phenomenon.

Cartwright (1979) proposes to meet the Simpson problem by imposing further conditions on the concept of one event being a cause of another. In particular, she wants to require that the increase in probability characteristic of prima facie causes defined above is considered only in situations that are "otherwise causally homogeneous with respect to" the effect. I am skeptical that we can know when situations are causally homogeneous. In the kind of example considered earlier concerning high intake of vitamin A being a potential inhibitor of lung cancer, it is certainly not possible to know or even to consider causally homogeneous situations. This is true of most applications of causal notions in nonexperimental settings and even in many experimental situations. I am also skeptical at a conceptual or philosophical level that we have any well-defined notion of homogeneity. Consider, for example, the data from Berkeley just described. There is no reason that we could not also pursue additional hypotheses. We might want to look at partial data from each department where the data were restricted just to the borderline cases. We might test the hypothesis that the female applicants were more able than the males but that at the borderline there was bias against the females. So far as I know, such a more refined analysis of the data has not been performed but there is no reason conceptually that we might not find something by entertaining such additional questions. My point is that there is no end to the analysis of data in a practical sense. We can, of course, exhaust finite data theoretically by considering all possible combinations, but this is only of mathematical significance.

A conflict of intuition can arise as to when to stop the refinement of data analysis. From a practical standpoint, many professional situations require detailed rules about such matters. The most obvious example is in the definition of classes for actuarial tables. What should be the variables relevant to fixing the rates on insurance policies? I have in mind here not only life insurance but also automobile insurance, property insurance, etc. I see a conflict at the most fundamental level between those who think there is some ultimate stopping point that can be determined in the analysis and those who do not.

There is another point to be mentioned about the Simpson problem. It is that if we can look at the data after they have been collected and if the probabilities in question are neither zero nor one, it is then easy to artificially define events that render any prima facie cause spurious. Of course, in ordinary statistical methodology it would be regarded as a scandal to construct such an event after looking at the data, but from a scientific standpoint the matter is not so simple. Certainly, looking at data that do not fit desired hypotheses or favorite theories is one of the best ways to get ideas about new hypotheses or new theories. But without further investigation we do not take seriously the ex post facto artificial construction of concepts. What is needed is another experiment or another set of data to determine whether the hypotheses in question are of serious interest. There is, however, another point to be made about such artificial concepts constructed solely by looking at the data and counting the outcomes. It is that somehow we need to exclude such concepts to avoid the undesirable outcome of every prima facie cause being spurious, at least every naturally hypothesized prima facie cause. One way to do this of course is to characterize the notion of genuine cause relative to a given set of concepts that may be used to define events considered as causes. Such an emendation and explicit restriction on the definition given above of genuine cause seems appropriate.[1]

2. MACROSCOPIC DETERMINISM

Even if one accepts the general argument that there is randomness in nature at the microscopic level, there continues to be a line of thought that in analysis of causality in ordinary experience it is useful and, in fact, in some cases almost mandatory to assume determinism. I will not try to summarize all the literature here but will concentrate on the arguments given in Hesslow (1976,1981), which attempt to give a deep-running argument against probabilistic causality, not just my particular version of it. (In addition to these articles of Hesslow, the reader is also referred to Rosen [1978] and for a particularly thorough critique of deterministic causality, Rosen [1982].)

[1] As Cartwright (1979) points out, it is a historical mistake to attribute Simpson's paradox to Simpson. The problem posed was already discussed in Cohen and Nagel's well-known textbook (1934), and according to Cartwright, Nagel believes that he learned about the problem from Yule's classic textbook of 1911. There has also been a substantial recent discussion of the paradox in the psychological literature (Hintzman 1980; Martin 1981).

As a formulation of determinism that avoids the global character of Laplace's, both Hesslow and Rosen cite Anscombe's (1975, p. 63) principle of relevant difference, "If an effect occurs in one case and a similar effect does not occur in an apparently similar case, then there must be a relevant further difference." Although statistical or probabilistic techniques are employed in testing hypotheses in the biological and social sciences, Hesslow claims that "there is nothing that shows that these hypotheses *themselves* are probabilistic in nature. In fact one can argue that the opposite is true, for statistics are commonly used in a way that presupposes determinism, namely, in various kinds of eliminative arguments."

Hesslow's intuitions here are very different from mine, so there is a basic conflict that could best be resolved by extensive review of the biological, medical, and social science literature. I shall not attempt that here but state what I think is wrong with one of Hesslow's ideal examples. He says that these kinds of eliminative arguments all have a simple structure. He takes the case of Jones, who had a fatal disease but was given a newly discovered medicine and recovered. We conclude, he says, that the cause of his recovery was M, the event of taking medicine. Now he says at the beginning that Jones had a "universally fatal disease." The first thing to challenge is the use of the adverb *universally*. This is not true of all the diseases of interest. Almost no diseases that are the subject for analysis and study by doctors are universally fatal. It is a familiar fact that when medicine is given we certainly like to attribute the recovery to medicine. But ordinarily the evidence is not overwhelming, because in the case of almost all diseases there is evidence of recovery of individuals who were not treated by the medicine. This is true of all kinds of diseases, from the plague to pneumonia. In making this statement, I am certainly not asserting that medicine is not without efficacy but only that Hesslow's claim is far too simple. The actual data do not support what he says.

Hesslow's claim that this is a case of determinism is puzzling because in his own explicit formulation of the argument he says, "Thus, (probably) M caused R," where R is the event of recovery. He himself explicitly introduces the caveat of probability. What he states is that "because something caused the recovery and, other causes apparently being scarce, M is the most likely candidate." Determinism comes in the use of *something*, but the conclusion he draws is probabilistic in character and could just as well have been drawn if he had started with the view that in most cases an identifiable agent caused the recovery but that in the remaining cases the recovery was spontaneous. Moreover, I would claim that there is no powerful argument for the determinism of the kind Hesslow was trying to give. One could look from one end of the medical literature to the other and simply not find the kind of need for the premises he talks about.

There is a point to be clear about on this matter. Because one is not endorsing determinism as a necessary way of life for biological and social scientists, it does not mean that the first identification of a probabilistic cause brings a scientific investigation of a given phenomenon to an end. It is a difficult and delicate matter to determine when no further causes can be identified. I am not offering any algorithms for making this determination. I am just making a strong claim that we do get along in practice with probabilistic results and we do not fill them out in an interesting deterministic fashion.

3. TYPES AND TOKENS

There are a host of conflicting intuitions about whether causality should mainly be discussed in terms of event types or event tokens, and also how the two levels are related. I restrict myself here to two issues, both of which are fundamental. One is whether cases of individual causation must inevitably be subsumable under general laws. The second is whether we can make inferences about individual causes when the general laws are merely probabilistic.

A good review of the first issue on subsumption of individual causal relations under general laws is given by Rosen (1982), and I shall not try to duplicate her excellent discussion of the many different views on this matter. Certainly, nowadays probably no one asserts the strong position that if a person holds that a singular causal statement is true then the person must hold that a certain appropriate covering law is true. One way out, perhaps most ably defended by Horgan (1980) is to admit that direct covering laws are not possible but that there are at work underneath precise laws, formulated in terms of precise properties that do give us the appropriate account in terms of general laws. But execution of this program certainly is at present, and in my own view will forever be, at best a pious hope. In many cases we shall not be able to supply the desired analysis.

There is a kind of psychological investigation that would throw interesting light on actual beliefs about these matters. Epistemological or philosophical arguments of the kind given by Horgan do not seem to me to be supportable. It would be enlightening to know if most people believe that there is such an underlying theory of events and if somehow it gives them comfort to believe that such a theory exists. The second and more particular psychological investigation would deal with the kinds of beliefs individuals hold and the responses they give to questions about individual causation. Is there a general tendency to subsume our causal accounts of individual events under proto-covering laws? It should be evident what I am saying about this first issue. The defense that there are laws either of a covering or a foundational nature cannot be defended on philosophical grounds, but it would be useful to transform the issue into a number of psychological questions as to what people actually do believe.

The second issue is in a way more surprising. It has mainly been emphasized by Hesslow. It is the claim that inferences from generic statistical relations to individual causal relations are necessarily invalid. Thus, he concludes that "if all generic causal relations are statistical, then we must either accept invalid inferences or refrain from talking about individual causation at all" (1981, p. 598). It seems to me that this line of argument is definitely mistaken and I would like to try to say why as clearly as I can. First of all, I agree that one does not make a logically or a mathematically valid argument from generic statistical relations to individual causal relations. It is in the nature of probability theory and its applications that the inference from the general to the particular is not in itself a mathematically valid inference The absence of such validity, however, in no way prohibits using generic causal relations that are clearly statistical in character to make inferences about individual causation. It is just that those inferences are not mathematically valid inferences—they are inferences made in the context of probability and uncertainty. I mention as an aside that there is a large literature by advocates of a relative frequency theory of probability about how to make inferences from relative frequencies to single cases. Since I come closer to being a Bayesian than a relative frequentist, I shall not review these arguments, but many of the discussions are relevant in arguing from a different viewpoint than mine about Hesslow's claims.

First, though, let me distinguish sharply between the generic relations and the individual relations and what I think is the appropriate terminology for making this distinction. The language I prefer is that the generic relations are average or mean relations. The individual relations at their fullest and best depend upon individual sample paths known in great detail. An individual sample path is the continuous temporal and spatial path of development of an individual's history. There is in this history ordinarily a great deal of information not available in simple mean data. I can say briefly and simply what the expected or mean life span is of an adult male who is now forty-five years old and is living in the United States, but if I consider some single individual and examine him in terms of his past history, his ancestors, his current state of health, his employment, etc., I may come to a very different view of his expected number of remaining years. Certainly it would be ludicrous to think that there is a logically valid inference from the mean data to the individual data.

But for a Bayesian or near Bayesian like myself, the matter has a rather straightforward solution. First of all, probabilities as matters of belief are directly given to individual events and their individual relationships. Second, by the standard theorem on total probability, when I say that a given individual has an expected lifetime of twenty years, I have already taken account of all the knowledge that I have about him. Of course, if I learn something new, the probability can change, just on the basis of the theorem on total probability. Now the central point is that ordinarily much of what I know about individuals is based upon generic causal relations. I simply do not know enough to go very much beyond generic relations, and thus my probabilistic estimate of an individual's expected remaining lifetime will very much depend on a few generic causal relations and not much else. The absence of logical validity in relating the generic to the individual in no way keeps me from talking about individual causation, contrary to Hesslow's claim. In fact, I would say that what I have said is just the right account of how we do talk about individual causation in the cases where we know something about generic probabilistic causal relations. We know, for example, that heavy clouds are a good sign of rain, and when accompanied by a drop in atmospheric pressure an even better sign. We know that these two conditions alone will not cause rain with probability one, but there is a strong probabilistic causal relation. We go on to say, well, rain is likely sometime this afternoon. We are quite happy with our causal views of the matter based on a couple of generic causal relations. Intimate details of the kind available to meteorologists with the professional responsibility to predict the weather are not available, let us say, in the instance being discussed. The meteorologist faced with a similar problem uses a much more complex theory of generic relations in order finally to issue his prediction for the afternoon. It is also important to note, of course, that on the kind of Bayesian view I am describing here there is no algorithm or simple calculus for passing by probability from generic causal relationships to individual ones, even for the trained meteorologist. It is a matter of judgment as to how the knowledge one has is used and assessed. The use of the theorem on total probability mentioned above depends on both conditional and unconditional probabilities, which in general depend on judgment. In the case where there is very fine scientific knowledge of the laws in question it might be on occasion that the conditional probabilities are known from extensive scientific experimentation, but then another aspect of the problem related to the application to the individual event will not be known from such scientific experimentation except in very unusual cases, and judgment will enter necessarily.

4. PHYSICAL FLOW OF CAUSES

In his excellent review article on probabilistic causality, Salmon (1980) puts his finger on one of the most important conflicting intuitions about causality. The derivations of the fundamental differential equations of classical physics give in most cases a very satisfying physical analysis of the flow of causes in a system, but there is no mention of probability. It is characteristic of the areas in which probabilistic analysis is used to a very large extent that a detailed theory of the phenomena in question is missing. The examples from medicine given earlier are typical. We may have some general ideas about how a vaccine works or about the mechanisms for absorbing vitamin A, but we do not have anything like an adequate detailed theory of these matters. We are presently very far from being able to make any kind of detailed theoretical predictions derived from fundamental assumptions about molecular structure, for example. Concerning these or related questions we have a very poor understanding in comparison with the kinds of models successful in various parts of classical physics about the detailed flow of causes. I think Salmon is quite right in pointing out that the absence of being able to give such an analysis is the source of the air of paradox of some of the counterexamples that have been given. The core argument is to challenge the claim that the occurrence of a cause should increase the probability of the occurrence of its effect.

Salmon uses as a good example of this phenomenon the hypothetical case made up by Deborah Rosen and reported in my 1970 monograph. A golfer makes a birdie by hitting a limb of a tree at just the right angle, not something that he planned to do. The disturbing aspect is that if we estimated the probability of his making a birdie prior to his making the shot and we added the condition that the ball hit the branch, we would ordinarily estimate the probability as being definitely lower than that he would have made a birdie without this given condition. On the other hand, when we see the event happen we have an immediate physical recognition that the exact angle that he hit the branch played a crucial role in the ball's going into the cup. In my 1970 discussion of this example, I did not take sufficient account of the conflict of intuition between the general probabilistic view and the highly structured physical view. I now think it is important to do so and I very much agree with Salmon that the issues here are central to a general acceptability of a probabilistic theory of causality. I therefore want to make a revised response.

There are at least three different kinds of cases in which what seem for other reasons to be prima facie causes in fact turn out to be negative causes, i.e., the conditional probability of the effect's occurring is lowered given the cause. One sort of case involves situations in which we know a great deal about the classical physics. A second kind of case is where an artificial example can be constructed and we may want to make claims about observing a causal chain. Salmon gives a succinct and useful example of this kind, which I discuss. Third, there are the cases in which we attribute without any grounds some surprising event as a cause of some significant effect. In certain respects the ancient predilection for omens falls under this category, but I shall not expand upon this view further.

In the first kind of case there is a natural description of the event after the fact that makes everything come out right. Using the golf ball example as typical, we now describe the event as that of the golf ball's hitting the branch at exactly the right angle to fall into the cup. Given such a description we would of course make the conditional probability close to one, but it is only after the fact that we could describe the event in this fashion. On the other hand, it is certainly too general to expect much to come out of the event described simply as the golf ball's hitting the limb of the tree. It is not really feasible to aim before the event at a detailed description of the event adequate to make a good physical prediction. We will not be given the values of parameters sufficiently precisely to predict that the golf ball will hit the limb of the tree at an angle just right for bouncing into the cup. Consequently, in such cases we cannot hope to predict the effects of such surprising causes, but based upon physical theories that are accurate to a high degree of approximation we understand that this is what happened after we have observed the sequence of events. Another way of putting the matter is that there is a whole range of cases in which we do not have much hope of applying in an interesting scientific or commonsense way probabilistic analysis, because the causes will be surprising. Even in cases of extraordinary conceptual simplicity, e.g., the N-body problem with only forces of gravitation acting between the bodies, extended prediction of behavior for any length of time is not in general possible. Thus, although a Bayesian in such matters, I confess to being unable to make good probabilistic causal analyses of many kinds of individual events. In the same fashion, I cannot apply to such events, in advance of their happening, detailed physical theories. The possibilities of application in both cases seem hopeless as a matter of prediction. This may not be the way we want the world to be but this is the way it is.

Salmon also gives an example that has a much simpler physical description than the golf ball example. It involves the eight ball and the cue ball on a pool table with the player having a 50–50 chance of sinking the eight ball with the cue ball when he tries. Moreover, the eight ball goes into the corner pocket, as Salmon says, "if and almost only if his cue ball goes into the other far corner pocket." Let event A be the player's attempting the shot, B the dropping of the eight ball in the corner pocket, and C the dropping of the cue ball into the other corner pocket. Under the hypotheses given, B is a prima facie cause of C, and Salmon is concerned about the fact that A does not screen B off from C, i.e., render B a spurious cause of C. Salmon expresses his concern by saying that we should have appropriate causal relations among A, B, and C without having to enter into more detailed physical theory. But it seems to me that this example illustrates a very widespread phenomenon. The physical analysis, which we regard as correct, namely, the genuine cause of C, i.e., the cue ball going into the pocket, is in terms of the impact forces and the direction of motion of the cue ball at the time of impact. We certainly believe that such specification can give us a detailed and correct account of the cue ball's motion. On the other hand, there is an important feature of this detailed physical analysis. We must engage in meticulous investigations; we are not able to make in a commonsense way the appropriate observations of these earlier events of motion and impact. In contrast, the events A, B, and C are obvious and directly observable. I do not find it surprising that we must go beyond these three events for a proper causal account, and yet at the same time we are not able to do so by the use of obvious commonsense events. Aristotle would not have had such an explanation, from all that we know about his physics. Why should we expect it of untutored common sense?

The second class of example, of which Salmon furnishes a very good instance, is when we know only probability transitions. The example he considers concerns an atom in an excited state. In particular, it is in the fourth energy level. The probability is one that it will necessarily decay to the zeroeth level, i.e., the ground state. The only question is whether the transitions will be through all the intermediate states three, two, and one, or whether some states will be jumped over. The probability of going directly from the fourth to the third state is 3/4 and from the fourth to the second state is 1/4. The probability of going from the third state to the first state is 3/4 and from the third state to the ground state 1/4. Finally, the probability of going from the second state to the first state is 1/4 and from the second state directly to the ground state 3/4. It is required also, of course, that the probability of going from the first state to the ground state is one. The paradox arises because of the fact that if a decaying atom occupies the second state in the process of decay, then the probability of its occupying the first state is 1/4, but the mean probability whatever the route taken of occupying the first state is the much higher probability of 10/16. Thus, on the probabilistic definitions given earlier of prima facie causes, occupying the second state is a negative prima facie cause of occupying the first state.

On the other hand, as Salmon emphasizes, after the events occur of the atom going from the fourth to the second to the first state, many would say that this sequence constitutes a causal chain. My own answer to this class of examples is to meet the problem head on and to deny that we want to call such sequences causal sequences. If all we know about the process is just the transition probabilities given, then occupancy of the second state remains a negative prima facie cause of occupying the first state. The fact of the actual sequence does not change this characterization. In my own constructive work on causality, I have not given a formal definition of causal chains, and for good reason. I think it is difficult to decide which of various conflicting intuitions should govern the definition.

We may also examine how our view of this example might change if the probabilities were made more extreme, i.e., if the mean probability of occupying the first energy state comes close to one and the probability of a transition from the second to the first state is close to zero. In such cases when we observe the sequence of transitions from the fourth to the second to the first state, we might be inclined to say that the atom decayed to the first state in spite of occupying the second state. By using such phrases as *in spite of* we indicate our skepticism that what we have observed is a genuine causal chain.

5. COMMON CAUSES

It was a virtue of Reichenbach to have recognized that a natural principle of causality is to expect events that are simultaneous, spatially separated, and strongly correlated, to depend upon some common cause to generate the correlation. There are a variety of controversial questions about the principle of common cause, and the source of the controversy is the absence of clear and widely accepted intuitions about what we should expect of such causes. Should we expect such causes to exist? Thus, when we observe phenomenologically simultaneous events strongly correlated, should we always be able to find a common cause that eliminates this phenomenological correlation in the sense that, when we condition on the common cause, the new conditional correlation is zero? Another question concerns the determinism of common causes. Ought we to expect such causes to be deterministic, or can we find common causes that are strictly probabilistic? In a recent essay, Van Fraassen (1982) expresses the view that the causes must be deterministic in the following way.

> But a belief in the principle of the common cause implies a belief that there is in the relevant cases not merely a compatibility (so that deterministic hidden variables could be intro-

> duced into models for the theory) but that all those hidden events which are the common causes, are real, and therefore, that the world is really deterministic (p. 208).

Salmon (1982) in his reply to Van Fraassen suggests that the principle of common cause is sometimes used as an explanatory principle and sometimes as a principle of inference. Also he implicitly suggests a third and different use as a maxim of rationality, which is a use also considered by Van Fraassen. The maxim is: search for a common cause whenever feasible to explain simultaneous events that are strongly correlated. Using the principle as a maxim does not guarantee any explanations nor any inferences but can be important in the strategy of research. The dialogue between Salmon and Van Fraassen in the two articles mentioned contains a number of useful points about common causes, but rather than consider in detail their examples, counterexamples, arguments, and counterarguments to each other, I want to suggest what I think is a reasonable view of the principle of common cause. In doing so I shall avoid references to quantum mechanics except in one instance. I shall also generalize the discussion to more than two events, because in many scientific applications it is not adequate to consider the correlations of only two events.

First let me say more explicitly what I shall mean by common cause. The exposition here will be rather sketchy. The technical details of many of the points made are given in the Appendix.

Let A and B be events that are approximately simultaneous and let

$$P(AB) \neq P(A)P(B);$$

i.e., A and B are not independent but correlated. Then the event C is a *common cause* of A and B if

(i) C occurs earlier than A and B;

(ii) $P(AB|C) = P(A|C)P(B|C)$;

(iii) $P(AB|\overline{C}) = P(A|\overline{C})P(B|\overline{C})$.

In other words, C renders A and B conditionally independent, and so does $\overline{C}$, the complement of C. When the correlation between A and B is positive, i.e., when

$$P(AB) > P(A)P(B),$$

we may also want to require:

(iv) C is a prima facie cause of A and of B.

I shall not assume (iv) in what follows. I state in informal language a number of propositions that are meant to clarify some of the controversy about common causes. The first two propositions follow from a theorem about common causes proved in Suppes and Zanotti, (1981).

PROPOSITION I. *Let events A_1, A_2, $\cdots$, A_n be given with any two of the events correlated. Then a necessary and sufficient condition for it to be possible to construct a common cause of these events is that the events $A_1, A_2, \ldots, A_n$ have a joint probability distribution compatible with the given pairwise correlations.*

An important point to emphasize about this proposition is its generality and at the same time its weakness. There are no restrictions placed on the nature of the common causes. Once any sorts of restrictions of a physical or other empirical kind are imposed, then the common cause might not exist. If we simply want to know whether a common cause can be found as a matter of principle as an underlying cause of the observed correlations between events, then the answer is not one that has been much discussed in the literature. All that is required is the existence of a joint probability distribution of the phenomenological variables. It is obvious that if the candidates for common causes are restricted in advance, then it is a simple matter to give artificial examples that show that among possible causes given in advance no common cause can be found. The ease with which such artificial examples are constructed makes it obvious that the same holds true in significant scientific investigations. When the possible causes of diseases are restricted, for example, it is often difficult for physicians to find a common cause among the given set of candidates.

PROPOSITION II. *The common cause of Proposition I can always be constructed so as to be deterministic.*

Again, without restriction, determinism is always open to us. On the other hand, it is easy to impose some natural principles of symmetry that exclude deterministic causes when the correlations are strictly probabilistic, i.e., the correlations between the events at the phenomenological level are not themselves deterministic. Explicit formulations of these principles of symmetry are given in the Appendix.

PROPOSITION III. *Conditions of symmetry can easily be found such that strictly probabilistic correlations between phenomenologically observed events have as a common cause one that is strictly probabilistic.*

This last proposition is special in nature, of course. It refers to principles of symmetry discussed in the Appendix. The conditions are sufficient but not necessary. It would be desirable to find significant necessary and sufficient conditions that require the common cause to be probabilistic rather than deterministic in character.

Finally, I state one application to quantum mechanics.

PROPOSITION IV. *There are correlated phenomenological data that cannot have a common cause that is theoretically consistent with quantum mechanics, because there can be no joint probability distribution of the data, as described in Proposition I.*

APPENDIX ON COMMON CAUSES

In this Appendix I present a number of theorems about inferences from phenomenological correlations to common causes. In the framework of quantum mechanics, the theorems are mainly theorems about hidden variables. Most of the proofs will not be given, but references will be cited where they may be found. The content of this Appendix follows closely the first part of Suppes and Zanotti (1984).

To emphasize conceptual matters and to keep technical simplicity in the forefront, I consider only two-valued random variables taking the values ±1. We shall also assume symmetry for these random variables in that their expectations will be zero and thus they will each have a positive variance of one. For emphasis we state:

GENERAL ASSUMPTION. *The phenomenological random variables* $\mathbf{X}_1, \ldots, \mathbf{X}_N$ *have possible values* ± 1, *with means* $E(\mathbf{X}_i) = 0, 1 \leq i \leq N$.

We also use the notation $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ for phenomenological random variables. We use the notation $E(\mathbf{XY})$ for covariance, which for these symmetric random variables is also the same as their correlation $\rho(\mathbf{X}, \mathbf{Y})$.

The basic meaning of *common cause* that we shall assume is that when two random variables, say $\mathbf{X}$ and $\mathbf{Y}$, are given, then in order for a hidden variable $\boldsymbol{\lambda}$ to be labeled a common cause, it must render the random variables conditionally independent, that is,

$$E(\mathbf{XY}|\boldsymbol{\lambda}) = E(\mathbf{X}|\boldsymbol{\lambda})E(\mathbf{Y}|\boldsymbol{\lambda}). \tag{1}$$

We begin with a theorem asserting a deterministic result. It says that if two random variables have a strictly negative correlation, then any cause in the sense of (1) must be deterministic, that is, the conditional variances of the two random variables, given the hidden variable $\boldsymbol{\lambda}$, must be zero. We use the notation $\sigma(\mathbf{X}|\boldsymbol{\lambda})$ for the conditional standard deviation of $\mathbf{X}$ given $\boldsymbol{\lambda}$, and its square is, of course, the conditional variance.

THEOREM 1. (Suppes and Zanotti, 1976). *If*

(i) $E(\mathbf{XY}|\boldsymbol{\lambda}) = E(\mathbf{X}|\boldsymbol{\lambda})E(\mathbf{Y}|\boldsymbol{\lambda})$

(ii) $\rho(\mathbf{X},\mathbf{Y}) = -1$

then

$$\sigma(\mathbf{X}|\boldsymbol{\lambda}) = \sigma(\mathbf{Y}|\boldsymbol{\lambda}) = 0.$$

The second theorem asserts that the only thing required to have a common cause for N random variables is that they have a joint probability distribution. This theorem is conceptually important in relation to the long history of hidden variable theorems in quantum mechanics. For example, in the original proof of Bell's inequalities, Bell (1964) assumed a causal hidden variable in the sense of (1) and derived from this assumption his inequalities. What Theorem 2 shows is that the assumption of a hidden variable is not necessary in such discussions—it is sufficient to remain at the phenomenological level. Once we know that there exists a joint probability distribution then there must be a causal hidden variable, and in fact this hidden variable may be constructed so as to be deterministic.

THEOREM 2. (Suppes and Zanotti, 1981). *Given phenomenological random variables* $\mathbf{X}_1,\ldots,\ \mathbf{X}_N$ *then there exists a hidden variable* $\boldsymbol{\lambda}$, *a common cause, such that*

$$E(\mathbf{X}_1,\ldots,\mathbf{X}_N|\boldsymbol{\lambda}) = E(\mathbf{X}_1|\boldsymbol{\lambda})\cdots E(\mathbf{X}_N|\boldsymbol{\lambda})$$

if and only if there exists a joint probability distribution of $\mathbf{X}_1,\ldots,\mathbf{X}_N$. *Moreover,* $\boldsymbol{\lambda}$ *may be constructed as a deterministic cause, i.e., for* $1 \leq i \leq N$

$$\sigma(\mathbf{X}_i|\boldsymbol{\lambda}) = 0.$$

6. EXCHANGEABILITY

We now turn to imposing some natural symmetry conditions both at a phenomenological and at a theoretical level. The main principle of symmetry we shall use is that of exchangeability. Two random variables $\mathbf{X}$ and $\mathbf{Y}$ of the class we are studying are said to be exchangeable if the following probabilistic equality is satisfied.

$$(2) \qquad P(\mathbf{X} = 1, \mathbf{Y} = -1) = P(\mathbf{X} = -1, \mathbf{Y} = 1).$$

The first theorem we state shows that if two random variables are exchangeable at the phenomenological level then there exists a hidden causal variable satisfying the additional restriction that they have the same conditional expectation if and only if their correlation is not negative.

THEOREM 3. (Suppes and Zanotti, 1980). *If* $\mathbf{X}$ *and* $\mathbf{Y}$ *are exchangeable, then there exists a hidden variable* $\boldsymbol{\lambda}$ *such that*

(i) $\boldsymbol{\lambda}$ *is a common cause of* $\mathbf{X}$ *and* $\mathbf{Y}$,

(ii) $E(\mathbf{X}|\boldsymbol{\lambda}) = E(\mathbf{Y}|\boldsymbol{\lambda})$

if and only if

$$\rho(\mathbf{X},\mathbf{Y}) \geq 0.$$

There are several remarks to be made about this theorem. First, the phenomenological principle of symmetry, namely, the principle of exchangeability, has not been used in physics as explicitly as one might expect. In the context of the kinds of experiments ordinarily used to test hidden variable theories, the requirement of phenomenological exchangeability is uncontroversial. On the other hand, the theoretical requirement of identity of conditional distributions does not have the same status. We emphasize that we refer here to the expected causal effect of $\boldsymbol{\lambda}$. Obviously the actual causal effects will in general be quite different. We certainly would concede that in many physical situations this principle may be too strong. The point of our theorems is to show that once such a strong theoretical principle of symmetry is required then exchangeable and negatively correlated random variables cannot satisfy it.

Theorem 4 strengthens Theorem 3 to show that when the correlations are strictly between zero and one then the common cause cannot be deterministic.

THEOREM 4. (Suppes and Zanotti, 1984). *Given the conditions of Theorem 3, if* $0 < \rho(\mathbf{X},\mathbf{Y}) < 1$ *then* $\boldsymbol{\lambda}$ *cannot be deterministic, i.e.,* $\sigma(\mathbf{X}|\boldsymbol{\lambda}), \sigma(\mathbf{Y}|\boldsymbol{\lambda}) \neq 0$.

Proof. We first observe that under the assumptions we have made:

$$\mathrm{Min}\{P(\mathbf{X}=1,\mathbf{Y}=-1), P(\mathbf{X}=1,\mathbf{Y}=1), P(\mathbf{X}=-1,\mathbf{Y}=-1)\} > 0.$$

Now, let Ω be the probability space on which all random variables are defined. Let $\mathcal{A} = \{A_i\}, 1 \leq i \leq N$ and $\mathcal{H} = \{H_j\}, 1 \leq j \leq M$ be two partitions of Ω. We say that $\mathcal{H}$ *is a refinement of* $\mathcal{A}$ *in probability* if and only if for all i's and j's we have:

$$\text{If } P(A_i \cap H_j) > 0 \text{ then } P(A_i \cap H_j) = P(H_j).$$

Now let $\boldsymbol{\lambda}$ be a causal random variable for $\mathbf{X}$ and $\mathbf{Y}$ in the sense of Theorem 3, and let $\boldsymbol{\lambda}$ have induced partition $\mathcal{H} = \{H_j\}$, which without loss of generality may be assumed finite. Then $\boldsymbol{\lambda}$ is deterministic if $\mathcal{H}$ is a refinement in probability of the partition $\mathcal{A}= \{A_i\}$ generated by $\mathbf{X}$ and $\mathbf{Y}$, for assume, by way of contradiction, that this is not the case. Then there must exist i and j such that $P(A_i \cap H_j) > 0$ and

$$P(A_i \cap H_j) < P(H_j),$$

but then $0 < P(A_i|H_j) < 1$.

We next show that if $\boldsymbol{\lambda}$ is deterministic then $E(\mathbf{X}|\boldsymbol{\lambda}) \neq E(\mathbf{Y}|\boldsymbol{\lambda})$, which will complete the proof.

Let, as before, $\mathcal{H} = \{H_j\}$ be the partition generated by $\boldsymbol{\lambda}$. Since we know that

$$\Sigma_j P(\mathbf{X} = 1, \mathbf{Y} = -1, H_j) = P(\mathbf{X} = 1, \mathbf{Y} = -1) > 0$$

there must be an H_j such that

$$P(\mathbf{X} = 1, \mathbf{Y} = -1, H_j) > 0,$$

but since $\boldsymbol{\lambda}$ is deterministic, $\mathcal{H}$ must be a refinement of $\mathcal{A}$ and thus as already proved

$$P(\mathbf{X} = 1, \mathbf{Y} = -1|H_j) = 1,$$

whence

$$\begin{aligned} P(\mathbf{X} = 1, \mathbf{Y} = 1|H_j) &= 0 \\ P(\mathbf{X} = -1, \mathbf{Y} = 1|H_j) &= 0 \\ P(\mathbf{X} = -1, \mathbf{Y} = -1|H_j) &= 0, \end{aligned}$$

and consequently we have

$$\left\{ \begin{array}{l} P(\mathbf{X} = 1|H_j) = P(\mathbf{Y} = -1|H_j) = 1 \\ P(\mathbf{X} = -1|H_j) = P(\mathbf{Y} = 1|H_j) = 0 \end{array} \right. \tag{3}$$

Remembering that $E(\mathbf{X}|\boldsymbol{\lambda})$ is a function of $\boldsymbol{\lambda}$ and thus of the partition $\mathcal{H}$, we have from (3) at once that

$$E(\mathbf{X}|\boldsymbol{\lambda}) \neq E(\mathbf{Y}|\boldsymbol{\lambda}).$$

11

WHEN ARE PROBABILISTIC EXPLANATIONS POSSIBLE?

The primary criterion of adequacy of a probabilistic causal analysis is that the causal variable should render the simultaneous phenomenological data conditionally independent. The intuition back of this idea is that the common cause of the phenomena should factor out the observed correlations. So we label the principle the *common cause criterion.* If we find that the barometric pressure and temperature are both dropping at the same time, we do not think of one as the cause of the other but look for a common dynamical cause within the physical theory of meteorology. If we find fever and headaches positively correlated, we look for a common disease as the source and do not consider one the cause of the other. But we do not want to suggest that satisfaction of this criterion is the end of the search for causes or probabilistic explanations. It does represent a significant and important milestone in any particular investigation.

Under another banner the search for common causes in quantum mechanics is the search for hidden variables. A hidden variable that satisfies the common cause criterion provides a satisfactory explanation "in classical terms" of the quantum phenomenon. Much of the earlier discussion of hidden variables in quantum mechanics has centered around the search for deterministic underlying processes, but for some time now the literature has also been concerned with the existence of probabilistic hidden vari-

*Reprinted from *Synthese*, **48** (1981), 191-199. Written jointly with Mario Zanotti.

ables. It is a striking and important fact that even probabilistic hidden variables do not always exist when certain intuitive criteria are imposed. One of the simplest examples was given by Bell in 1971, who extended his earlier deterministic work to construct an inequality that is a consequence of assuming that two pairs of values of experimental settings in spin-1/2 experiments must violate a necessary consequence of the common cause criterion, that is, the requirement that a hidden variable render the data conditionally independent. It is easy to show that Bell's inequality is a necessary but not sufficient condition for conditional independence. However, we shall not pursue further matters involving specific quantum mechanical phenomena in the present context.

Our aims in this short article are more general. First we establish a necessary and sufficient condition for satisfaction of the common cause criterion for events or two-valued random variables. The condition is existence of a joint probability distribution. We then consider the more difficult problem of finding necessary and sufficient conditions for the existence of a joint distribution. We state and prove a general result only for the case of three (two-valued) random variables, but it has as a corollary a pair of new Bell-type inequalities.

The limitation from a scientific standpoint of the first result on satisfaction of the common cause criterion is evident. The mere theoretical existence of a common cause is often of no interest. The point of the theorem is clarification of the general framework of probabilistic analysis. The theorem was partially anticipated by some unpublished work of Arthur Fine on deterministic hidden variables.

The second theorem about the existence of a joint distribution is more directly applicable as a general requirement on data structures, for it is easy to give examples of three random variables for which there can be no joint distribution. Consider the following. Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be two-valued random variables taking the values 1 and -1. Moreover, let us restrict the expectation of the three random variables to being zero, that is,

$$E(\mathbf{X}) = E(\mathbf{Y}) = E(\mathbf{Z}) = 0.$$

Now assume that the correlation of $\mathbf{X}$ and $\mathbf{Y}$ is -1, the correlation of $\mathbf{Y}$ and $\mathbf{Z}$ is -1, and the correlation of $\mathbf{X}$ and $\mathbf{Z}$ is -1. It is easy to show that there can be no joint distribution of these three random variables.

THEOREM ON COMMON CAUSES. *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *be two-valued random variables. Then a necessary and sufficient condition that there is a random variable* $\boldsymbol{\lambda}$ *such that* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *are conditionally independent given* $\boldsymbol{\lambda}$ *is that there exists a joint probability distribution of* $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

Proof. The necessity is trivial. By hypothesis

$$P(\mathbf{X}_1 = 1, \ldots, \mathbf{X}_n = 1 | \boldsymbol{\lambda} = \lambda) = \Pi_{i=1}^{n} P(\mathbf{X}_i = 1 | \boldsymbol{\lambda} = \lambda).$$

We now integrate with respect to $\boldsymbol{\lambda}$, which has, let us say, measure μ, so we obtain

$$P(\mathbf{X}_1 = 1, \ldots, \mathbf{X}_n = 1) = \int P(\mathbf{X}_1 = 1, \ldots, \mathbf{X}_n = 1 | \boldsymbol{\lambda} = \lambda) = d\mu(\lambda).$$

The argument for sufficiency is more complex. To begin with, let Ω be the space on which the joint distribution of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is defined. Each $\mathbf{X}_i$ generates a partition of Ω:

$$A_i = \{\omega : \omega \in \Omega \ \& \ \mathbf{X}_i(\omega) = 1\}$$

$$\overline{A}_i = \{\omega : \omega \in \Omega \ \& \ \mathbf{X}_i(\omega) = -1\}.$$

Let $\mathcal{P}$ be the partition that is the common refinement of all these two-element partitions, i.e.,

$$\mathcal{P} = \{A_1 \ldots A_n, A_1 \ldots \overline{A}_n, \ldots, \overline{A}_i \ldots \overline{A}_n\},$$

where juxtaposition denotes intersection. Obviously $\mathcal{P}$ has 2^n elements. For brevity of notation we shall denote the elements of partition $\mathcal{P}$ by C_j, and the indicator function for C_j by $\widehat{\mathbf{C}}_j$, i.e.,

$$\widehat{\mathbf{C}}_j(\omega) = \begin{cases} 1 \text{ if } \omega \in C_j \\ 0 \text{ otherwise.} \end{cases}$$

We now define the desired random variable $\boldsymbol{\lambda}$ in terms of the $\mathbf{C}_j$.

$$\boldsymbol{\lambda} = \sum \alpha_j \widehat{\mathbf{C}}_j \tag{1}$$

where the α_j are distinct real numbers, i.e., $\alpha_i \neq \alpha_j$ for $i \neq j$. The distribution μ of $\boldsymbol{\lambda}$ is obviously determined by the joint distribution of the random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

Using (1), we can now express the conditional expectation of each $\mathbf{X}_i$ and of their product given $\boldsymbol{\lambda}$.

$$E(\mathbf{X}_i | \boldsymbol{\lambda}) = \textstyle\sum_j \frac{\widehat{C}_j}{P(C_j)} \int_{C_j} \mathbf{X}_i d\mu(\boldsymbol{\lambda}) \tag{2}$$

and

$$E(\mathbf{X}_1 \cdots \mathbf{X}_n | \boldsymbol{\lambda}) = \textstyle\sum_j \frac{\widehat{C}_j}{P(C_j)} \int_{C_j} \mathbf{X}_1 \cdots \mathbf{X}_n d\mu(\boldsymbol{\lambda}). \tag{3}$$

We need to show that the product of (2) over the $\mathbf{X}_i$'s is equal to (3). We first note that in the case of (2) or (3) the integrand, $\mathbf{X}_i$ in one case, the product $\mathbf{X}_1 \cdots \mathbf{X}_n$ in the other, has value 1 or -1. (So $\boldsymbol{\lambda}$ as constructed is deterministic—a point we comment on later.) Second, the integral over the region C_j is just $P(C_j)$. So we have

$$E(\mathbf{X}_i|\boldsymbol{\lambda}) = \sum_j sgn_{C_j}(\mathbf{X}_i)\widehat{\mathbf{C}}_j \tag{4}$$

where $sgn_{C_j}(\mathbf{X}_i)$ is 1 or -1, as the case may be for $\mathbf{X}_i$ over the region C_j. From (4) we then have

$$\Pi_{i=1}^n E(\mathbf{X}_i|\lambda) = \Pi_i \sum_j sgn_{C_j}(\mathbf{X}_i)\widehat{\mathbf{C}}_j. \tag{5}$$

Given that the product $\widehat{\mathbf{C}}_j\widehat{\mathbf{C}}_{j'} = 0$, if $j \neq j'$, we may interchange product and summation in (5) to obtain

$$\Pi_i E(\mathbf{X}_{\mathbf{i}}|\boldsymbol{\lambda}) = \sum_j \Pi_i sgn_{C_j}(\mathbf{X}_i)\widehat{C}_j, \tag{6}$$

but by the argument already given the right-hand side of (6) is equal to $E(\mathbf{X}_1 \cdots \mathbf{X}_n|\boldsymbol{\lambda})$ as desired.

There are several comments we want to make about this theorem and its proof. First, because the random variables $\mathbf{X}_i$ are two-valued, it is sufficient just to consider their expectations in analyzing their conditional independence. Second, and more important, the random variable $\boldsymbol{\lambda}$ constructed in terms of the partition $\mathcal{P}$ yields a deterministic solution. This may be satisfying to some, but it is important to emphasize that the artificial character of $\boldsymbol{\lambda}$ severely limits its scientific interest. What the theorem does show is that the general structural problem of finding a common cause of a finite collection of events or two-valued random variables has a positive abstract solution. Moreover, extensions to infinite collections of events or continuous random variables are possible but the technical details will not be entered into here. We do emphasize that the necessary inference from conditional independence to a joint distribution does not assume a deterministic causal structure.

The place where the abstract consideration of common causes has been pursued the most vigorously is, of course, in the analysis of the possibility of hidden variables in quantum mechanics. Given the negative results of Bell already mentioned, it is clear how the Theorem on Common Causes must apply: the phenomenological events in question do not have a joint distribution. We are reserving for another occasion the detailed consideration of this point.

Within the present general framework it is important to explore further the existence of nondeterministic common causes. Many important constructive examples of such causes are to be found in many parts of science, but the general theory needs more development. One simple example is given at the end of this article.

We turn now to the second theorem about the existence of a joint distribution for three two-valued random variables, which could be the indicator functions, for example, for three events. We assume the possible values as 1 and -1, and the expectations are zero, so the variances are 1 and the covariances are identical to the correlations.

JOINT DISTRIBUTION THEOREM. *Let* $\mathbf{X}$, $\mathbf{Y}$, *and* $\mathbf{Z}$ *be random variables with possible values* 1 *and* -1, *and with*

$$E(\mathbf{X}) = E(\mathbf{Y}) = E(\mathbf{Z}) = 0.$$

Then a necessary and sufficient condition for the existence of a joint probability distribution of the three random variables is that the following two inequalities be satisfied.

$$-1 \leq E(\mathbf{XY}) + E(\mathbf{YZ}) + E(\mathbf{XZ}) \leq 1 + 2\ \mathrm{Min}\{E(\mathbf{XY}), E(\mathbf{YZ}), E(\mathbf{XZ})\}.$$

Proof. We first observe that

$$(1) \qquad E(\mathbf{XY}) = p_{11\cdot} - p_{10\cdot} - p_{01\cdot} + p_{00\cdot},$$

where

$$p_{10\cdot} = P(\mathbf{X} = 1, \mathbf{Y} = -1), \text{etc.}$$

(We use 0 rather than -1 as a subscript for the -1 value for simplicity of notation. The dot refers to $\mathbf{Z}$.)
It follows easily from (1) that

$$(2) \qquad p_{00\cdot} = p_{11\cdot} = \tfrac{1}{4} + \frac{E(\mathbf{XY})}{4},$$

and similarly

$$(3) \qquad p_{0\cdot0} = p_{1\cdot1} = \tfrac{1}{4} + \frac{E(\mathbf{XZ})}{4},$$

$$(4) \qquad p_{\cdot00} = p_{\cdot11} = \tfrac{1}{4} + \frac{E(\mathbf{YZ})}{4},$$

$$(5) \qquad p_{01\cdot} = p_{10\cdot} = \tfrac{1}{4} - \frac{E(\mathbf{XY})}{4},$$

$$(6) \qquad p_{0\cdot1} = p_{1\cdot0} = \tfrac{1}{4} - \frac{E(\mathbf{XZ})}{4},$$

$$(7) \qquad p_{\cdot01} = p_{\cdot10} = \tfrac{1}{4} - \frac{E(\mathbf{YZ})}{4}.$$

Using (2)–(7) we can directly derive the following seven equations for the joint distribution—with p_{111} being treated as a parameter along with $E(\mathbf{XY})$, $E(\mathbf{YZ})$, and $E(\mathbf{XZ})$:

$$\left\{\begin{aligned}
p_{110} &= \tfrac{1}{4} + \tfrac{E(\mathbf{XY})}{4} - p_{111}\\
p_{101} &= \tfrac{1}{4} + \tfrac{E(\mathbf{XZ})}{4} - p_{111}\\
p_{011} &= \tfrac{1}{4} + \tfrac{E(\mathbf{YZ})}{4} - p_{111}\\
p_{100} &= p_{111} - \tfrac{E(\mathbf{XY})}{4} - \tfrac{E(\mathbf{XZ})}{4}\\
p_{010} &= p_{111} - \tfrac{E(\mathbf{XY})}{4} - \tfrac{E(\mathbf{YZ})}{4}\\
p_{001} &= p_{111} - \tfrac{E(\mathbf{XZ})}{4} - \tfrac{E(\mathbf{YZ})}{4}\\
p_{000} &= \tfrac{1}{4} + \tfrac{E(\mathbf{XY})}{4} + \tfrac{E(\mathbf{XZ})}{4} + \tfrac{E(\mathbf{YZ})}{4} - p_{111}
\end{aligned}\right. \tag{8}$$

From (8) we derive the following inequalities, where $\alpha = 4p_{111}$:

$$\left\{\begin{aligned}
&1 + E(\mathbf{XY}) \geq \alpha\\
&1 + E(\mathbf{XZ}) \geq \alpha\\
&1 + E(\mathbf{YZ}) \geq \alpha\\
&E(\mathbf{YZ}) + E(\mathbf{XZ}) \leq \alpha\\
&E(\mathbf{XY}) + E(\mathbf{YZ}) \leq \alpha\\
&E(\mathbf{YZ}) + E(\mathbf{XZ}) \leq \alpha\\
&1 + E(\mathbf{XY}) + E(\mathbf{XZ}) + (\mathbf{YZ}) \geq \alpha
\end{aligned}\right. \tag{9}$$

From the last inequality of (9), we have at once

$$-1 \leq E(\mathbf{XY}) + E(\mathbf{XZ}) + (\mathbf{YZ}), \tag{10}$$

because α must be nonnegative. Second, taking the maximum of the fourth, fifth, and sixth inequalities and the minimum of the first, second, and third, and adding $\mathrm{Min}(E(\mathbf{XY}), E(\mathbf{XZ}), E(\mathbf{YZ}))$ to both sides, we obtain

$$E(\mathbf{XY}) + E(\mathbf{XZ}) + (\mathbf{YZ}) \leq 1 + 2\ \mathrm{Min}\{E(\mathbf{XY}), E(\mathbf{XZ}), E(\mathbf{YZ})\}. \tag{11}$$

Inequalities (10) and (11) represent the desired result. Their necessity, i.e., that they must hold for any joint distribution of $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, is apparent from their derivation.

Sufficiency follows from the following argument. Let

$$\begin{aligned} C_1 &= \mathrm{Max}\{E(\mathbf{XY})+E(\mathbf{XZ}), E(\mathbf{XY})+E(\mathbf{YZ}), E(\mathbf{XZ})+E(\mathbf{YZ})\}, \\ C_2 &= \mathrm{Min}\{E(\mathbf{XY}), E(\mathbf{XZ}), E(\mathbf{YZ})\}. \end{aligned}$$

It is an immediate consequence of (10) and (11) that

$$(12) \qquad C_1 \leq 1 + C_2,$$

$$(13) \qquad 1 + C_1 + C_2 \geq 0.$$

Assume now that $C_1 \geq 0$.
We may then choose $\alpha = 4p_{111}$ so that

$$\alpha = \beta C_1 + (1-\beta)(1+C_2), \quad \text{for } 0 \leq \beta \leq 1.$$

On the other hand, if $C_1 < 0$, choose α so that

$$\alpha = \beta(1 + C_1 + C_2), \quad \text{for } 0 \leq \beta \leq 1.$$

It is straightforward to show that for either case of C_1, any choice of β in the closed interval [0,1] will define an $\alpha/4 = p_{111}$ satisfying the distribution equation (8).

The two theorems we have proved can be combined to give a pair of Bell-type inequalities. Two differences from Bell's 1971 results are significant. First, we give not simply necessary, but necessary and sufficient conditions for existence of a hidden variable. Second, we deal with three rather than four random variables. As would be expected from the proofs of the two theorems, our method of attack is quite different from Bell's.

The corollary is an immediate consequence of the two theorems.

COROLLARY ON HIDDEN VARIABLES. *Let* $\mathbf{X}$, $\mathbf{Y}$, *and* $\mathbf{Z}$ *be random variables with possible values* 1 *and* -1, *and with*

$$E(\mathbf{X}) = E(\mathbf{Y}) = E(\mathbf{Z}) = 0.$$

Then a necessary and sufficient condition for the existence of a hidden variable or common cause $\boldsymbol{\lambda}$ *with respect to which the three given random variables are conditionally independent is that the phenomenological correlations satisfy the inequalities*

$$-1 \leq E(\mathbf{XY}) + E(\mathbf{YZ}) + E(\mathbf{XZ}) \leq 1 + 2\ \mathrm{Min}\{E(\mathbf{XY}), E(\mathbf{YZ}), E(\mathbf{XZ})\}.$$

NONDETERMINISTIC EXAMPLE. The deterministic result of the Theorem on Common Causes can, as already indicated, be misleading. We conclude with a simple but important example that is strictly probabilistic.

Let $\mathbf{X}$ and $\mathbf{Y}$ be two random variables that have a bivariate normal distribution with $|\rho(\mathbf{X}, \mathbf{Y}| \neq 1$, i.e., the correlation to be factored out by a common cause is nondeterministic, and without loss of generality $E(\mathbf{X}) = E(\mathbf{Y}) = 0$. It is a standard result that the partial correlation of $\mathbf{X}$ and $\mathbf{Y}$ with $\mathbf{Z}$ held constant is (for a proof, see Suppes, 1970, p. 116).

$$\rho(\mathbf{XY} \cdot \mathbf{Z}) = \frac{\rho(\mathbf{X}, \mathbf{Y}) - \rho(\mathbf{X}, \mathbf{Z})\rho(\mathbf{Y}, \mathbf{Z})}{\sqrt{1 - \rho^2(\mathbf{X}, \mathbf{Z})}\sqrt{1 - \rho^2(\mathbf{Y}, \mathbf{Z})}}.$$

Because a multivariate normal distribution is invariant under an affine transformation, we may take

$$E(\mathbf{Z}) = 0,$$
$$E(\mathbf{Z}^2) = 1.$$

If $\rho(\mathbf{X}, \mathbf{Y}) \geq 0$, we set

$$\rho(\mathbf{X}, \mathbf{Z}) = \rho(\mathbf{Y}, \mathbf{Z}) = \sqrt{\rho(\mathbf{X}, \mathbf{Y})}.$$

If $\rho(\mathbf{X}, \mathbf{Y}) < 0$, we set

$$\rho(\mathbf{X}, \mathbf{Z}) = -\rho(\mathbf{Y}, \mathbf{Z}) = \sqrt{|\rho(\mathbf{X}, \mathbf{Y})|}.$$

It is straightforward to check that we now have a proper multivariate normal distribution of $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ with

$$\rho(\mathbf{XY} \cdot \mathbf{Z}) = 0$$

and $\rho(\mathbf{X}, \mathbf{Z})$ and $\rho(\mathbf{Y}, \mathbf{Z})$ nondeterministic.

12

NON-MARKOVIAN CAUSALITY IN THE SOCIAL SCIENCES WITH SOME THEOREMS ON TRANSITIVITY

When we consider familiar observable properties of a physical object we postulate almost without thinking that these properties are determined by the particular atomic structure of the object at the moment of observation. If we know the current atomic structure we firmly believe that it is not necessary to know anything about the history of the object. It may well be that in many practical instances this assumption is a theoretical one that we cannot put into practice, but it is a deep and important theoretical assumption about the Markovian character of the physical world. It is a standard theoretical move in physics to postulate a concept of state such that if we know the state of a system at a given time we need know nothing about the system at any earlier time in order to analyze and predict its future behavior. This radical Markovian truncation of the past is one of the most essential general concepts in the physical sciences.

It is an important methodological and scientific question to what extent a similar viewpoint can be made to work in the social sciences. I have deliberately not said that it was a general philosophical issue. The reason

*Reprinted from *Synthese*, **68** (1986), 129-140.

for this omission is obvious. It is reasonable to believe that a person's beliefs and actions at a given time are determined by the current encoding of his past experience in his central nervous system and by the current state of the many chemical substances in his body at the present instant, together with the current circumstances of his environment. Almost none of us accept a philosophical view of action at a distance across time so that an event that occurred in the past directly affects an action taking place now. In the present instance, however, our philosophical views although perhaps correct in general principle are too complacent and do not readily form a basis in many areas for serious scientific endeavor. The difficulty is easy to describe. We are not able to give a theory or description of the current state of a person, or more generally of a society, with sufficient accuracy and detail to be of much direct use in scientific analysis of any personal or social phenomenon of interest.

The scientific problem is that of being able to postulate detailed internal states that have essential properties of uniformity across many different situations. The great success of the physical sciences has depended upon the structural identity of substances, at least in relation to the phenomenological properties we have as yet investigated with any thoroughness. In essential ways, all atoms of a given kind, for example, hydrogen, mercury, etc., are identical in structure, or there are in almost all circumstances a very small number of variants. In contrast, it seems a plausible negative thesis that in the case of persons nothing like such uniformity of structure holds for the properties we consider essential, for example, the internal psychological structure of a person's memory, feelings, etc. There is much to support such a negative thesis at the present time and, therefore, many reasons to be skeptical that a powerful and scientifically useful concept of state can be introduced in ways that render the postulated processes of a person or a social group Markovian in character.

Whatever the status of the general conceptual argument I have been trying to give, the empirical evidence on the kinds of models that are actually used in the social sciences very much supports my thesis. What I want to do in the remainder of this article is to explore various aspects of this non-Markovian kind of analysis, and to speculate on its consequences for theory construction.

To begin with, I show that transitivity of probabilistic causality does not depend upon a Markov condition although, as has been shown by Eells and Sober (1983), such a Markov condition is sufficient even though too restrictive. It is easy to want to hold that any reasonable theory of causality should be transitive in character, that is, if A is a cause of B and B is a cause of C, then A should be a cause of C. As we shall see, this is a characteristic feature of a wide class of non-Markovian processes. It would be disturbing for the theory of non-Markovian causality if this were not so. In the second section, I in fact turn to examples of such processes, drawn from psychology, in particular from learning theory. The third section considers examples of such processes familiar from econometrics.

To avoid any misunderstanding in the examples that follow, I note that by *Markov condition* I mean first-order Markov condition. We can, of course, have a second-order or a third-order Markov condition especially in the case of discrete trials or time periods, but the basic intuition about states, and the familiar use of the concept in physics, is certainly only in terms of first-order Markov processes. In some of the examples, only a finite segment of the past is included, but that is usually for purposes of practical simplification. Further refinements and more accurate analysis can be obtained by further extension into the past.

1. TRANSITIVITY OF NON-MARKOVIAN CAUSES

In Suppes (1970) I gave a specific counterexample to show that probabilistic causation need not be transitive. For the purposes of this discussion I have in mind my definition of prima facie cause. Event $B_{t'}$ is a *prima facie cause* of event A_t if and only if (i) $t' < t$—these subscripts refer, of course, to time of occurrence of an event, (ii) $P(B_{t'}) > 0$, and (iii) $P(A_t|B_{t'}) > P(A_t)$. Eells and Sober (1983), as already mentioned, showed that a Markov condition is sufficient to guarantee transitivity of prima facie causes, as just defined. To facilitate comparison with their proof I use the same letters for causes as they do: $C_{t''}$ is a prima facie cause of $F_{t'}$, which is itself a prima facie cause of E_t, with $t'' < t' < t$. Hereafter I drop the time subscript since it plays no role in the proof. As for other notation, juxtaposition of letters standing for events denotes intersection and a bar over such a letter denotes complementation.

THEOREM 1. *Given*

(i) $P(F|C) > P(F|\overline{C})$,
(ii) $P(E|F) > P(E|\overline{F})$,
(iii) $P(E|FC) \geq P(E|F\overline{C})$ *and* $P(E|\overline{F}C) \geq P(E|\overline{F}\overline{C})$,

with $P(FC), P(F\overline{C}), P(\overline{F}C), P(\overline{F}\overline{C}) > 0$.

Then

$$P(E|C) > P(E|\overline{C}).$$

Before giving the proof, note that condition (iii) replaces the Markov condition. Cause C can directly affect the occurrence of E (strict inequality in (iii)). Equality in both conjuncts of (iii) is just the first-order Markov condition. I label (iii) the principle of *remote causes*, for in the natural temporal interpretation, C is remote in time from E rather than contiguous. The name does not say as much as it should, for it is also important that C act on E in the same positive way that it acts on F.

Proof. We have at once from the theorem on total probability

$$P(E|C) = P(E|FC)P(F|C) + P(E|\overline{F}C)P(\overline{F}|C) \tag{1}$$

and

$$P(E|\overline{C}) = P(E|F\overline{C})P(F|\overline{C}) + P(E|\overline{F}\overline{C})P(\overline{F}|\overline{C}). \tag{2}$$

From (i) and (ii)

$$P(E|F)(P(F|C) - P(F|\overline{C})) + P(E|\overline{F})(P(F|\overline{C}) - P(F|C)) > 0. \tag{3}$$

Using now $P(\overline{F}|C) = 1 - P(F|C)$ and $P(\overline{F}|\overline{C}) = 1 - P(F|\overline{C})$, we obtain from (3)

$$P(E|F)(P(F|C) - P(F|\overline{C})) + P(E|\overline{F})(P(\overline{F}|C) - P(\overline{F}|\overline{C})) > 0, \tag{4}$$

and so rearranging (4)

$$P(E|F)P(F|C) + P(E|\overline{F})P(\overline{F}|C) > P(E|F)P(F|\overline{C}) + P(E|\overline{F})\, P(\overline{F}|\overline{C}). \tag{5}$$

It follows easily from (iii) that

$$P(E|FC) \geq P(E|F) \geq P(E|F\overline{C}) \tag{6}$$

and

$$P(E|\overline{F}C) \geq P(E|\overline{F}) \geq P(E\overline{F}\overline{C}). \tag{7}$$

From (5)–(7) we infer

$$P(E|FC)P(F|C) + P(E|\overline{F}C)P(\overline{F}|C) > P(E|F\overline{C})P(F|\overline{C}) + P(E|\overline{F}\overline{C})P(\overline{F}|\overline{C}). \tag{8}$$

From (1), (2), and (8) we have at once, as desired,

$$P(E|C) > P(E|\overline{C}).$$

It is immediately obvious that the counterexample to transitivity given earlier by me (Suppes 1970, p. 58) does not satisfy condition (iii). It is also easy to show that (iii) is sufficient but not necessary for transitivity.

Eells and Sober (1983) also prove a theorem involving several intermediary causes. The event C causes each of the events F_i and they in turn cause E. Their Markov condition in this case can be generalized in a similar fashion—see (III) of Theorem 2 below. For comparison, I again use their notation. Let $F_1, \ldots, F_n$ be the intermediate causes; then each $K_i, i = 1, \ldots, 2^n$, is the intersection of all the F_j's or their complements. Eells and Sober call the K_i's *specifications*. The notation K_{ji} is for the specification of the F's excepting F_j. Then for each j, $i = 1, \ldots, 2^{n-1}$.

THEOREM 2. *If*

$$\begin{array}{ll} \text{(I) } P(F_j|C) > P(F_j|\overline{C}), & \text{for } j = 1, \ldots, n \\ \text{(II) } P(E|K_{ji}F_j) > P(E|K_{ji}\overline{F}_j), & \text{for } j = 1, \ldots, n, \\ & \quad i = 1, \ldots, 2^{n-1} \\ \text{(III) } P(E|K_iC) \geq P(E|K_i\overline{C}), & \text{for } i = 1, \ldots, 2^n \end{array}$$

with $P(K_iC), P(K_i\overline{C}) > 0$,

(IV) *The F_j's are mutually independent conditional on C and also conditional on $\overline{C}$,*

then

$$P(E|C) > P(E|\overline{C}).$$

Proof. The proof consists of showing that (III), together with (I), (II), and (IV), imply the Markov case already established by Eells and Sober. More explicitly, let (III′) be:

$$\text{(III}') \qquad P(E|K_iC) = P(E|K_i\overline{C}) \text{ for } i = 1, \ldots, 2^n.$$

Eells and Sober prove that (I), (II), (III′), and (IV) imply the conclusion of Theorem 2.

The important step in the reduction is to prove

$$(1) \qquad P(E|K_iC) \geq P(E|K_i) \geq P(E|K_i\overline{C}), i = 1, \ldots, 2^n.$$

but, as was already seen in similar fashion in the proof of Theorem 1, we have for each i:

$$(2) \qquad P(E|K_i) = P(E|K_iC)P(C|K_i) + P(E|K_i\overline{C})P(\overline{C}|K_i),$$

but since $P(C|K_i), P(\overline{C}|K_i) \geq 0$ and $P(C|K_i) + P(\overline{C}|K_i) = 1$ we infer (1) from (2) and (III).

Now let

$$(3) \qquad \begin{aligned} a_i &= P(F_j|C) \\ \overline{a}_i &= P(\overline{F}_j|C) \\ b_i &= P(F_j|\overline{C}) \\ \overline{b}_i &= P(\overline{F}_j|\overline{C}), \text{ for } j = 1, \ldots, n. \end{aligned}$$

And let

$$(4)\qquad \begin{aligned} s_i &= P(E|K_iC) \\ \bar{s}_i &= P(E|K_i\overline{C}), \text{ for } i = 1,\ldots,2^n. \end{aligned}$$

Then, with the obvious extensions of this notation, we have:

$$(5)\qquad P(E|C) = a_1 \cdots a_n s_1 + a_1 \cdots \bar{a}_n s_2 + \cdots + \bar{a}_1 \cdots \bar{a}_n s_{2^n},$$

and

$$(6)\qquad P(E|\overline{C}) = b_1 \cdots b_n \bar{s}_1 + b_1 \cdots \bar{b}_n \bar{s}_2 + \cdots + \bar{b}_1 \cdots \bar{b}_n \bar{s}_{2^n},$$

where $K_1 = F_1 \cdots F_n, K_2 = F_1 \cdots \overline{F}_n, \ldots, K_{2^n} = \overline{F}_1 \cdots \overline{F}_n$.

Now define

$$(7)\qquad s_i^* = \bar{s}_i^* = P(E|K_i).$$

Define $P^*(E|C)$ to be $P(E|C)$ with each s_i replaced by s_i^*, and define similarly $P^*(E|\overline{C})$. Thus

$$(8)\qquad P^*(E|C) = a_1 \cdots a_n s_1^* + \cdots + \bar{a}_1 \cdots \bar{a}_n s_{2^n}^*$$

and

$$(9)\qquad P^*(E|\overline{C}) = b_1 \cdots b_n \bar{s}_1^* + \cdots + \bar{b}_1 \cdots \bar{b}_n \bar{s}_{2^n}^*.$$

First, in view of (1), we have at once

$$(10)\qquad P(E|C) \geq P^*(E|C),$$

since all terms of $P(E|C)$ are nonnegative. Similarly

$$(11)\qquad P^*(E|\overline{C}) \geq P(E|\overline{C}).$$

But Eells and Sober's theorem is just:

$$(12)\qquad P^*(E|C) > P^*(E|\overline{C}),$$

where the Markov assumption is incorporated in (8) and (9).

From (10)–(12) we have at once the desired result:

$$(13)\qquad P(E|C) > P(E|\overline{C}).$$

2. NON-MARKOVIAN LEARNING MODELS

To illustrate various specific points about causality, I shall draw on the theory of linear learning models already used for similar purposes in my 1970 monograph. Although I shall not dwell on the many applications of these models, they have been used extensively in psychology and more recently in control theory, especially by a number of Russian mathematicians and scientists.

For simplicity, let us assume that on every trial the organism can make exactly one of two responses, A_1 or A_2, and after each response it receives a reinforcement, E_1 or E_2, of one of the two possible responses. A learning parameter θ, which is a real number such that $0 < \theta \preceq 1$, describes the rate of learning in a manner to be made definite in a moment. A possible realization of the theory is an ordered triple $X = (X, P, \theta)$ of the following sort. X is the set of all sequences or ordered pairs such that the first member of each pair is an element of some set A and the second member an element of some set B, where A and B each have two elements. Intuitively, the set A represents the two possible responses and the set B the two possible reinforcements. P is a probability measure on the σ-algebra of cylinder sets of X, and θ is a real number as already described. To define the models of the theory, we need a certain amount of notation. Let $A_{i,n}$ be the event of response A_i on trial n; $E_{j,n}$ the event of reinforcemcnt E on trial n, where $i, j = 1, 2$; and for x in X let x_n be the equivalence class of all sequences in X which are identical with x through trial n. We may then characterize the theory by the following set-theoretical definition.

DEFINITION. *A triple* $\mathcal{X} = (X, P, \theta)$ *is a* linear learning model *if and only if the following two axioms are satisfied:*

A1 *If* $P(E_{i,n}A_{i',n}x_{n-1}) > 0$ *then*
$P(A_{i,n+1}|E_{i,n}A_{i',n}x_{n-1}) = (1-\theta)P(A_{i,n}|x_{n-1}) + \theta$;

A2 *If* $P(E_{j,n}A_{i',n}x_{n-1}) > 0$ *and* $i \neq j$ *then*
$P(A_{i,n+1}|E_{j,n}A_{i',n}x_{n-1}) = (1-\theta)P(A_{i,n}|x_{n-1})$.

As is clear from the two axioms, this linear response theory is intuitively very simple. The first axiom just says that when a response is reinforced, the probability of making that response on the next trial is increased by a simple linear transformation. The second axiom says that if some other response is reinforced, the probability of making the response is decreased by a second linear transformation.

The theoretical models of the theory of linear learning are determined by three types of parameters. First, a numerical value for the learning parameter θ must be selected; second, the initial probability of an A_1

response must be selected, that is, the probability $P(A_{1,1})$; and third, a reinforcement schedule must be chosen.

To illustrate various ideas, let us pick a Markov reinforcement schedule with $P(E_{1,1}) = \gamma$, and

$$\begin{array}{c|cc} n \diagdown n+1 & E_1 & E_2 \\ \hline E_1 & \alpha & 1-\alpha \\ E_2 & 1-\beta & \beta \end{array}$$

Also, let $P(A_{1,1}) = p_1$.

Let $\alpha = \gamma = p_1 = 0.5$, and let $\beta = 0.25$. Then it is easy to show

$$P(E_{1,2}|E_{2,1}) > P(E_{1,2}|E_{1,1})$$
$$P(A_{1,3}|E_{1,2}) > P(A_{1,3}|E_{2,2}),$$

but transitivity fails, for

$$P(A_{1,3}|E_{2,1}) < P(A_{1,3}|E_{1,1}).$$

It is easy to see that condition (iii) of Theorem 1 fails, for

$$P(A_{1,3}|E_{1,2}E_{2,1}) < P(A_{1,3}|E_{1,2}E_{1,1}).$$

In this example, the Markov reinforcement schedule upsets the expected transitivity, but there is a certain additional oddity present. In most applications of this kind of theory we do not tend to think in terms of one reinforcement causing another. Note that if we change parameters α and β to: $\alpha = \beta = 0.75$, then we have the transitive chain: $E_{1,1}$ is a prima facie cause of $E_{1,2}, E_{1,2}$ is a prima facie cause of $A_{1,3}$, and also, $E_{1,1}$ is a prima facie cause of $A_{1,3}$.

A more interesting point about causality in the linear learning model as formulated is that if we consider the entire temporal sequence of reinforcements preceding a response, then the preceding responses are all rendered causally irrelevant, even though for many reinforcement schedules $A_{i,n-1}$ is a prima facie cause of $A_{i,n}$, for $i = 1, 2$. On the other hand, it is easy to generalize the theory so that this causal irrelevance is removed. The change in probability of a response following a reinforcement of that response also depends on whether that response actually occurred just before the reinforcement. This generalization also has a body of experiments to support it. Once it is made, preceding responses as well as reinforcements are genuine causes within the theory, as so modified.

The linear learning models I have been discussing are examples of *chains of infinite order*, so called because the dependence on the past does not terminate after some fixed time or fixed number of trials. Still we find it very natural for a variety of reasons to think in terms of Markov processes, with the present state absorbing all needed information about the past, and so a definite effort has been made to redefine the concept of state for such chains of infinite order as linear learning models. When the reinforcement occurring on trial n is probabilistically dependent at most on the immediately preceding response on that trial, then the response probabilities can be taken as the states, and it is easy to show—under the restriction stated—that the process is Markov. Extensive examples of this Markovian approach are developed, as well as the general theory, in Norman (1972). However, even for the simple Markov reinforcement schedule introduced above, this approach will not work.

A more fundamental approach to "reducing" a linear learning model to a Markov one is to enlarge the set of psychological concepts. This reduction has been done by Estes and Suppes (1959b) by introducing concepts of stimulus sampling theory. The past history of the organism is rendered superfluous when the current state of conditioning is given. This state gives the conditioning relation of each stimulus to each response. The history of the learning that led to the current conditioning state is irrelevant to predictions of future responses.

For a certain class of phenomena such stimulus sampling models, which can also be described in terms of sampling hypotheses and having strategies with no change in the mathematics, can lead to a very satisfactory view of psychological phenomena. Unfortunately not even most cognitive experiments involving learning, let alone real-world learning, can be adequately dealt with by stimulus-sampling theory. Detailed analysis reverts to chains of infinite order. Correspondingly the systematic study of psychological causes in the real world must use, in a wide variety of cases, a non-Markovian setup.

3. ECONOMETRICS

Some of the most thoroughly studied empirical cases of non-Markovian causality are to be found in econometrics. Most weather forecasts are not made on the basis of a single set of simultaneous observations but from a set of observations extended in time. This is even more the case in economics where the causal data used in an analysis ordinarily extend back considerably further in time.

To illustrate the ideas, especially in a way that relates to the previous section, I examine an analysis of the causes of the levels of consumption in terms of individual disposable income in a given population. In this sort of analysis there is no claim that the amount of disposable income is the only cause of the level of consumption of individuals or households, but almost everyone would expect it to be a principal cause. The interesting question is whether present consumption is influenced by past disposable income. We also might expect such past influence would be stronger in the case of self-employed persons than in the case of wage-earners, although we shall not explore this idea.

To examine some models, let

c_{it} = consumption in time period t by individual or household i
d_{it} = disposable income of i in period t,

and for aggregation of n individuals, let

$$c_t = \sum_{i=1}^{n} c_{it}$$

$$D_t = \sum_{i=1}^{n} d_{it}.$$

(For a good discussion of such aggregation and other methodological aspects of the models considered in this section, see Malinvaud (1966, Ch. 4).)

Then an obvious linear model is:

$$(1) \qquad C_t = a_0 \sum_{\tau=1}^{T} b(t-\tau) D_{t-\tau} + e + \epsilon_t,$$

where a_0 and e are constants, $b(t-\tau)$ is a constant for period $t-\tau$ and ϵ_t is the error term for period t. In the usual probabilistic formulation it is assumed that the expectation of ϵ_t is zero. A natural specialization of (1) is to assume that the influence of disposable income on current consumption fades exponentially with time, and so we set

$$(2) \qquad b(t-\tau) = b^{\tau}.$$

If we take the period of t and τ to be one year, a variety of empirical studies show that the influence of the past as reflected in the linear model of (1), or (1) augmented by (2), is significant.

For example, a classic study by Friedman (1957) of annual data on consumption and disposable income for heads of household in the United States for the period 1905-1951 but excluding the war years yields the following numerical version of (1):

$$C_t = 0.29D_t + 0.19D_{t-1} + 0.13D_{t-2} + 0.09D_{t-3} + 0.06D_{t-4} + 0.04D_{t-5} - 4. \tag{2}$$

According to (2) about 30% of an increase in income would be used for additional consumption in the current year, but it would be five years before more than 90% was so used. Defense of the validity of (2) is not my concern. The point in the present context is to show that workaday empirical economics is often non-Markovian in character, and no effective method of changing the situation seems even remotely in sight.

The issue of validity here is not, however, purely academic. A traditional view of consumers is that changes in real income are quickly translated into changes of consumption. Following out this "quick-adaptation" assumption, it is argued that changes in income brought about by tax changes are a significant countercyclical force for stabilization of the economy. The other main view, reflected in the data of (2), is the life cycle/permanent income hypothesis. This hypothesis is that consumers slowly alter their consumption with changes of income and the rate of change depends on their perception of the extent to which the income change is temporary. Extensions of (2) and substantial microeconomic household data bearing on the alternative hypothesis are to be found in Hall and Mishkin (1982).

I also note that extensive recent theoretical discussion of causality in econometrics by Granger (1969), Sims (1972), and others is in a non-Markovian framework. More explicitly, let $(\mathbf{X}_t, \mathbf{Y}_t)$ be a stochastic process with $t = \cdots - 1, 0, 1, \ldots$ Granger defines "$\mathbf{Y}$ does not cause $\mathbf{X}$" as: the (minimum mean square error) linear predictor of $\mathbf{X}_{t+1}$ based on $\mathbf{X}_t, \mathbf{X}_{t-1}, \ldots, \mathbf{Y}_t, \mathbf{Y}_{t-1}, \ldots$ is identical to the linear predictor based on the $\mathbf{X}$-process alone, i.e., $\mathbf{X}_t, \mathbf{X}_{t-1}, \ldots$.

Philosophical views of causality—at least if it is intended for them to be relevant to theoretical and empirical work in the social sciences—must not be restricted to the dominant Markovian conceptions of causality that have played such a central role in physics.

PART III

PROBABILITY AND MEASUREMENT

13

FINITE EQUAL-INTERVAL MEASUREMENT STRUCTURES

In this article I consider some of the simplest non-trivial examples of measurement structures. The basic sets of objects or stimuli will in all cases be finite, and the adequacy of the elementary axioms for various structures depends heavily on this finiteness.

In addition to their finiteness, the distinguishing characteristic of the structures considered is that the objects are equally spaced in an appropriate sense along the continuum, so to speak, of the property being measured. The restrictions of finiteness and equal spacing enormously simplify the mathematics of measurement, but it is fortunately not the case that the simplification is accompanied by a total separation from realistic empirical applications. Finiteness and equal spacing are characteristic properties of many standard scales, for example, the ordinary ruler, the set of standard weights used with an equal-arm balance in the laboratory or shop, or almost any of the familiar gauges for measuring pressure, temperature, or volume.

Four kinds of such structures are dealt with, and each of them corresponds to a more general set of structures analyzed in the comprehensive treatise of Krantz, Luce, Suppes and Tversky (1971). The four kinds of structures are for extensive, difference, bisection, and conjoint measurement.

*Reprinted from *Theoria*, **38** (1972), 45-63.

1. EXTENSIVE MEASUREMENT

The distinction between extensive and intensive properties or magnitudes is a very old one in the history of science and philosophy. Extensive magnitudes are ones that can be added; e.g., mass and length are extensive magnitudes or quantities. Intensive magnitudes, in contrast, cannot be added, even though they can be measured. Two volumes of gas, e.g., with the same temperature, do not combine to form a gas with twice the temperature. It has been claimed by some theorists, e.g., Campbell (1920, 1928), that fundamental measurement of intensive magnitudes is not possible. However, I do not find the negative arguments of Campbell and others at all persuasive, and many examples of measurement structures provide a concrete refutation of Campbell's thesis.

I develop the axioms of extensive measurement in this section with three specific interpretations in mind. One is for the measurement of mass on an equal-arm balance, one is for the measurement of length of rigid rods, and one is for the measurement of subjective probabilities. Other interpretations are certainly possible, but I shall restrict detailed remarks to these three.

From a formal standpoint the basic structures are triples $\langle X, \mathcal{F}, \succeq \rangle$ where X is a non-empty set, $\mathcal{F}$ is a family of subsets of X and the relation $\succeq$ is a binary relation on $\mathcal{F}$. By using subsets of X as objects, we avoid the need for a separate primitive concept of concatenation. As a general structural condition, it shall be required that $\mathcal{F}$ be an *algebra of sets* on X, which is just to require that $\mathcal{F}$ be non-empty and be closed under union and complementation of sets, i.e., if A and B are in $\mathcal{F}$ then $A \cup B$ and $-A$ are also in $\mathcal{F}$.

The intended interpretations of the primitive concepts for the three cases mentioned is fairly obvious. In the case of mass, X is a set of physical objects, and for two subsets A and B, $A \succeq B$ if and only if the set A of objects is judged at least as heavy as the set B. It is probably worth emphasizing that several different uses of the equal-arm balance are appropriate for reaching a judgment of comparison. For example, if $A = \{x, y\}$ and $B = \{x, z\}$ it will not be possible literally to put A on one pan of the balance and simultaneously B on the other, because the object x is a member of both sets, but we can make the comparison in at least two different ways. One is just to compare the non-overlapping parts of the two subsets, which in the present case just comes down to the comparison of $\{y\}$ and $\{z\}$. A rather different empirical procedure that even eliminates the need for the balance to be equal arm is to first just balance A with sand on the other pan (or possibly water; but in either case, sand or water in small containers), and then to compare B

with this fixed amount of sand. Given the standard meaning of the set-theoretical operations of intersection, union, and complementation, no additional interpretations of these operations is required, even of union of sets, which serves as the operation of concatenation.

In the case of the rigid rods, the set X is just the collection of rods, and $A \succeq B$ if and only if the set A of rods, when laid end to end in a straight line, is judged longer than the set B of rods also so laid out. Variations on exactly how this qualitative comparison of length is to be made can easily be supplied.

In the case of subjective probabilities, the set X is the set of possible outcomes of the experiment or empirical situation being considered. The subsets of X in $\mathcal{F}$ are just events in the ordinary sense of probability concepts, and $A \succeq B$ if and only if A is judged at least as probable as B.

Axioms for extensive measurement, subject to the two restrictions of finitude and equal spacing, are given in the following definition. In the definition and subsequently we use the standard definitions for equivalence $\sim$ in terms of a weak ordering and also of a strict ordering. The definitions are just these: $A \sim B$ if and only if $A \succeq B$ and $B \succeq A$; $A \succ B$ if and only if $A \succeq B$, and not $B \succeq A$.

DEFINITION 1. *A structure* $\mathcal{X} = \langle X, \mathcal{F}, \succeq \rangle$ *is a* finite, equally spaced extensive structure *if and only if* X *is a finite set,* $\mathcal{F}$ *is an algebra of sets on* X*, and the following axioms are satisfied for every* A*,* B*, and* C *in* $\mathcal{F}$:

1. *The relation* $\succeq$ *is a weak ordering of* $\mathcal{F}$;
2. *If* $A \cap C = \emptyset$ *and* $B \cap C = \emptyset$*, then* $A \succeq B$ *if and only if* $A \cup C \succeq B \cup C$;
3. $A \succeq \emptyset$;
4. *Not* $\emptyset \succeq X$;
5. *If* $A \succeq B$ *then there is a* C *in* $\mathcal{F}$ *such that* $A \sim B \cup C$.

From the standpoint of the standard ideas about the measurement of mass or length, it would be natural to strengthen Axiom 3 to assert that if $A \neq \emptyset$, then $A \succ \emptyset$, but because this is not required for the representation theorem and is unduly restrictive in the case of subjective probabilities, the weaker axiom seems more appropriate.

In stating the representation and uniqueness theorem, we use the notion of an additive measure μ from $\mathcal{F}$ to the real numbers, i.e., a function μ such that for any A and B in $\mathcal{F}$

i. $\mu(\emptyset) = 0$,
ii. $\mu(A) \geq 0$,
iii. if $A \cap B = \emptyset$ then $\mu(A \cup B) = \mu(A) + \mu(B)$.

THEOREM 1. *Let $\mathcal{X} = \langle X, \mathcal{F}, \succeq \rangle$ be a finite, equally spaced extensive structure. Then there exists an additive measure μ such that for every A and B in $\mathcal{F}$*

$$\mu(A) \geq \mu(B) \text{ if and only if } A \succeq B.$$

The measure μ is unique up to a positive similarity transformation. Moreover, there are at most two equivalence classes of atomic events in $\mathcal{F}$; if there are two, one of these contains the empty event.

The proof of this theorem and much of the preceding discussion is to be found in Suppes (1969a, pp. 4–8).

2. DIFFERENCE MEASUREMENT

Referring to the distinction between extensive and intensive properties discussed at the beginning of the previous section, I could easily make a case for entitling this section *intensive measurement*, for it is characteristic of difference measurement that no operation corresponding to addition is present, and no meaningful combination of objects or stimuli is postulated for the difference structures.

In this section I shall deal with quaternary structures. As before, the basic set will be non-empty and finite, but in this case the relation on the set will be a quaternary relation. I will denote the basic set of objects by A and the quaternary relation by $\succeq$. The idea behind the quaternary relation $\succeq$ is that $ab \succeq cd$ holds when and only when the subjective (algebraic) difference between a and b is at least as great as that between c and d. In the case of similarity judgments, for example, the relation $\succeq$ would hold when the subject of an experiment judged that the similarity between a and b was at least as great as the similarity between c and d, due account being taken of the algebraic sign of the difference. The inclusion of the algebraic difference requires some care in interpretation; for example, in many similarity experiments a natural algebraic sign is not attached to the similarity. Instances that satisfy the present requirement are judgments of utility or of pitch or of intensity of sound; in fact, any kind of judgments in which the subject will recognize and accept that the judgments naturally lie along a one-dimensional continuum.

We define for the quaternary relation $\succeq$ just as for a binary relation, $\succ$ and $\sim$:

$ab \succ cd$ if and only if not $cd \succeq ab$,

$ab \sim cd$ if and only if $ab \succeq cd$ and $cd \succeq ab$.

It is also convenient to have at hand certain elementary definitions of the binary relation of strict precedence or preference and the relation $\sim$ of indifference or indistinguishability. These definitions are the following.

DEFINITION 2. *$a \succ b$ if and only if $ab \succ aa$.*

DEFINITION 3. *$a \sim b$ if and only if $ab \sim ba$.*

In order to express the equal-spacing part of our assumptions, we need one additional definition, namely, the definition that requires that adjacent objects in the ordering be equally spaced. For this purpose we introduce the definition of the binary relation J. The binary relation J is just the relation of immediate predecessor. Axiom 4 given below relates J to the quaternary relation $\sim$. The intuitive idea of Axiom 4 is just that if a stands in the relation J to b, and c stands in the relation J to d, then the difference between a and b is judged to be the same as the difference between c and d, due account being taken of algebraic sign.

DEFINITION 4. *aJb if and only if $a \succ b$ and for all c in A if $a \succ c$, then either $b \sim c$ or $b \succ c$.*

I now turn to the definition of finite equal-difference systems. The axioms given follow those given by Suppes and Zinnes (1963).

DEFINITION 5. *A quaternary structure $\mathfrak{A} = \langle A, \succeq \rangle$ is a* finite, equally spaced difference structure *if and only if the following axioms are satisfied for every a, b, c, and d in A:*

1. *The relation $\succeq$ is a weak ordering of $A \times A$;*
2. *If $ab \succeq cd$, then $ac \succeq bd$;*
3. *If $ab \succeq cd$, then $dc \succeq ba$;*
4. *If aJb and cJd, then $ab \sim cd$.*

Keeping in mind the empirical interpretations mentioned already, it is easy to grasp the intuitive interpretation of each axiom. The first axiom just requires that the quaternary relation $\succeq$ be a weak ordering in terms of the qualitative difference between objects or stimuli. Axiom 2 is the most powerful and fundamental axiom in many ways. It expresses a simple necessary property of the intended interpretation of the relation $\succeq$. Axiom 3 just expresses a necessary algebraic fact about the differences. Notice that Axioms 1–3 are necessary axioms. Only Axiom 4 is sufficient but not necessary; it expresses the equal-spacing assumption already discussed.

From these four axioms we can prove the following representation and uniqueness theorem.

THEOREM 2. *Let* $\mathfrak{A} = \langle A, \succeq \rangle$ *be a finite, equally spaced difference structure. Then there exists a real-valued function* φ *on* A *such that for every* a, b, c, *and* d *in* A

$$\varphi(a) - \varphi(b) \geq \varphi(c) - \varphi(d) \quad \textit{if and only if} \quad ab \succeq cd.$$

Moreover, if φ' *is any other real-valued function having the same property, then* φ *and* φ' *are related by a (positive) linear transformation, i.e., there exist real numbers* α *and* β *with* $\alpha > 0$ *such that for every* a *in* A

$$\varphi'(a) = \alpha\varphi(a) + \beta.$$

The proof of this theorem is given at the end of the article. In addition, a number of elementary properties are organized in a series of elementary lemmas leading up to the proof of the theorem.

Upon casual inspection it might be thought that the first three axioms of Definition 5 would characterize all finite-difference structures for which a numerical representation could be found. However, Scott and Suppes (1958) showed that the theory of all representable finite difference structures is not characterized by these three axioms and indeed cannot be characterized by any simple finite list of axioms.

It might be thought that with the addition of the non-necessary Axiom 4 it would be difficult to satisfy the axioms, because an arbitrary collection of stimuli or objects would not. However, if the stimuli being studied lie on a continuum, then it will be possible to select a standard sequence that will satisfy the axioms, just as is done in the case of selecting a standard set of weights for use on an equal-arm balance.

3. BISECTION MEASUREMENT

Relational structures closely related to the finite difference structures are bisection systems $\mathfrak{A} = \langle A, B \rangle$ where B is a ternary relation on the finite set A with the interpretation that $B(a, b, c)$ if and only if b is the midpoint of the interval between a and c. The method of bisection has a long history in psychophysics, but it is important to emphasize that satisfaction of the axioms given below requires no assumptions of an underlying physical measurement. All we need is the intuitive idea of a qualitative continuum, and even that is not needed for formal purposes. It is, of course, interesting, after the fundamental psychological measurement in terms of the method of bisection has been made, to construct a psychophysical function relating physical measurements of the same magnitude to psychological measurements. The axioms given below for the method of bisection imply a number of checks that should be satisfied before it is asserted that a numerical representing function exists, but these checks have often been ignored in the experimental literature that reports use of the method of bisection.

For the simplest set of axioms and definitions, we take both the bisection relation B and the ordering relation $\succeq$ as primitive, but it is easy to eliminate $\succeq$ by definition. We use the binary relation J as defined in the previous section (Definition 4).

DEFINITION 6. *A structure* $\mathfrak{A} = \langle A, \succeq, B\rangle$ *is a* bisection structure *if and only if the following axioms are satisfied for every* a, a', b, c, *and* c' *in* A:

1. *The relation* $\succeq$ *is a weak ordering of* A;
2. *If* $B(abc)$ *and* $B(abc')$ *then* $c \sim c'$;
3. *If* $B(abc)$ *and* $B(a'bc)$ *then* $a \sim a'$;
4. *If* $B(abc)$ *then* $a \succ b$ *and* $b \succ c$;
5. *If* aJb *and* bJc *then* $B(abc)$;
6. *If* $B(abc)$ *and* $a'Ja$ *and* cJc' *then* $B(a'bc')$.

The intuitive interpretation of the axioms is relatively transparent. The first axiom is already familiar. Axioms 2 and 3 require uniqueness of the endpoints up to equivalence, which clearly separates bisection from betweenness. Axiom 4 relates the ternary bisection relation and the binary ordering relation in a natural way, although it imposes a formal constraint on the bisection relation which would often be omitted. Inclusion of this order property as part of the relation B simplifies the axioms. Axiom 5 is a strong assumption of equal spacing, and Axiom 6 expresses an additional feature of this equal spacing. In view of the axioms given earlier for difference structures, it is somewhat surprising that Axiom 6 can be shown to be independent of Axiom 5, but it is easy to give a model of Axioms 1-5 to show that this is the case. For we can take a model with

$$B(abc) \text{ if and only if } aJb \text{ and } bJc$$

and satisfy all of the first five axioms.

The representation and uniqueness theorem assumes the following form.

THEOREM 3. *Let* $\mathfrak{A} = \langle A, \succeq, B\rangle$ *be a (finite) bisection structure. Then there exists a real-valued function* φ *defined on* A *such that for every* a, b, *and* c *in* A

(i) $\varphi(a) \geq \varphi(b)$ *if and only if* $a \succeq b$,

(ii) $2\varphi(b) = \varphi(a)+\varphi(b)$ *and* $\varphi(a) > \varphi(b) > \varphi(c)$ *if and only if* $B(a,b,c)$.

Moreover, any other real-valued function φ' *satisfying* (*i*) *and* (*ii*) *is related to* φ *by a* (*positive*) *linear transformation, i.e., there exist real numbers* α *and* β *with* $\alpha > 0$ *such that for all* a *in* A

$$\varphi'(a) = \alpha\varphi(a) + \beta.$$

The proof of this theorem is given in the final section.

4. CONJOINT MEASUREMENT

In many kinds of experimental or observational environments, it turns out to be the case that the measurement of a single magnitude or property is not feasible or theoretically interesting. What is of interest is the joint measurement of several properties simultaneously. In this section we consider axioms for additive *conjoint* measurement. The intended representation here is that we consider ordered pairs of objects or stimuli. The first members of the pairs are drawn from one set and consequently represent one kind of property or magnitude, and the second members of the pairs are objects drawn from a second set representing a different magnitude or property. Given the ordered-pair structure, we shall only require judgments of whether or not one pair jointly has more of the "conjoined" attribute than a second pair.

It is easy to give examples of interpretations for which this way of looking at ordered pairs is natural. Suppose we are asked to judge the capabilities of individuals to assume a position of leadership in an organization. What we are given about the individuals is their intelligence scores on an ordinal scale and a charisma measure on an ordinal scale. Thus for each individual we can say how he compares on each scale with any other individual. The problem is to make judgments as between the individuals in terms of their overall capabilities. The axioms given below indicate the kind of conditions that are sufficient to guarantee finite equally spaced conjoint measurement, where in this case the equal spacing is along each dimension.

As a second example, a pair (a,p) can represent a tone with intensity a and frequency p, and the problem is to judge which of two tones sounds louder. Thus the subject judges $(a,p) \succeq (b,q)$ if and only if tone (a,p) seems at least as loud as (b,q). Other examples from disciplines as widely separated as economics and physics are easily given, and are discussed in considerable detail in Krantz, Luce, Suppes and Tversky (1971, Ch. 6).

It is to be stressed that the additive representation sought in this section is a special case. Generalizations of additivity are discussed in the reference just cited. It is also to be noted that the restriction in this section to ordered pairs rather than ordered n-tuples is not essential.

Before turning to the axioms of (additive) conjoint measurement, we need a couple of elementary definitions that permit us to define ordering relations on the individual components. On the basis of the axioms on the ordering relation between pairs, we shall be able to prove that these ordering relations on the components are also weak orderings. In the following elementary definitions A_1 is the set of first components and A_2 the set of second components. Thus, when reference is made to an ordered pair (a,p), it is understood that a is in A_1 and p is in A_2.

DEFINITION 7. *$a \succeq b$ if and only if for all p in $A_2, (a,p) \succeq (b,p)$.*

In terms of this relation we define $a \succ b$ and $a \sim b$ in the usual fashion. Also, a similar definition is needed for the second component .

DEFINITION 8. *$p \succeq q$ if and only if for all a in A_1, $(a, p) \succeq (a,q)$.*

We also use the notation already introduced for the relation $\succeq$ on $A_1 \times A_2$, namely,

$$(a,p) \succ (b,q) \text{ if and only if not } (b,q) \succeq (a,p),$$

and

$$(a,p) \sim (b,q) \text{ if and only if } (a,p) \succeq (b,q) \text{ and } (b,q) \succeq (a,p).$$

Our axioms for additive conjoint measurement in the finite, equal-spacing case are embodied in the following definition.

DEFINITION 9. *A structure $\langle A_1, A_2, \geq \rangle$ is a* finite, equally spaced additive conjoint structure *if and only if the following axioms are satisfied for every a and b in A_1 and every p and q in A_2:*

1. *The relation $\succeq$ is a weak ordering on $A_1 \times A_2$;*
2. *If $(a,p) \succeq (b,p)$ then $(a,q) \succeq (b,q)$;*
3. *If $(a,p) \succeq (a,q)$ then $(b,p) \succeq (b,q)$;*
4. *If aJb and pJq then $(a,q) \sim (b,p)$.*

The intuitive content of the four axioms of Definition 9 is apparent, but requires some discussion. Axiom 1, of course, is the familiar requirement of a weak ordering. Axioms 2 and 3 express an independence condition of one component from the other. Thus Axiom 2 says that if the pair (a,p) is at least as great as the pair (b,p) then the same relationship holds when p is replaced by any other member q of A_2, and Axiom 3 says the same thing about the second component. Axiom 4 is, of course, sufficient but not necessary. It states the equal-spacing assumption, and corresponds closely to the corresponding axiom for finite, equally spaced difference structures.

It might be thought the monotonicity assumption that if $(a,p) \sim (b,q)$ and $a \succ b$, then $q \succ p$, also needs to be assumed as an axiom, but as we show in the proof of the representation theorem in the final section, this additional assumption is not necessary: it can be proved from the first four axioms alone.

The statement of the representation and uniqueness theorem, to which we now turn, assumes exactly the expected form. The only thing to note is that the two real-valued functions on each component are welded together by the same unit as reflected by the common change of unit α in the theorem, but a different origin is permitted.

THEOREM 4. *Let $\langle A_1, A_2, \succeq\rangle$ be a finite, equally spaced additive conjoint structure. Then there exist real-valued functions φ_1 and φ_2 on A_1 and A_2 respectively such that for a and b in A_1 and p and q in A_2*

$$\varphi_1(a) + \varphi_2(q) \geq \varphi_1(b) + \varphi_2(p) \textit{ if and only if } (a,q) \succeq (b,p).$$

Moreover, if φ_1' and φ_2' are any two other functions with the same property, then there exist real numbers α, α', β and γ with α, $\alpha' > 0$ such that

$$\varphi_1' = \alpha\varphi + \beta \textit{ and } \varphi_2' = \alpha' \; \varphi_2 + \gamma,$$

and if A_1 and A_2 each have at least two elements not equivalent in order, then $\alpha = \alpha'$.

It is worth noting that the uniqueness part of Theorem 4 has a natural geometrical interpretation. If we think of the functions φ_1 and φ_2 mapping pairs into the Cartesian plane, then the uniqueness theorem says that in the standard geometrical sense, any change of scale must be uniform in every direction, but the origin can be translated by a different distance along the different axes.

5. PROOFS

Proof of Theorem 2. Although the following elementary lemmas are not necessary to give a proof of Theorem 2, they are needed in a completely explicit discussion, and their inclusion will perhaps be useful in organizing the reader's thinking about difference structures, which are not as familiar as extensive structures. Indications of the proofs of the elementary lemmas are given only in a few instances.

All of the lemmas refer to a fixed quaternary structure $\mathfrak{A} = \langle A, \succeq\rangle$, and the binary relations $\succ$, $\sim$, and J defined in Section 2.

LEMMA 1. *The relation $\succ$ is asymmetric and transitive on A.*

LEMMA 2. *The relation $\sim$ is reflexive, symmetric, and transitive on A.*

LEMMA 3. *Exactly one of the following holds for any a and b in A: $a \succ b$, $b \succ a$, $a \sim b$.*

LEMMA 4. *If aJ^nb, then $a \succ b$.* (The proofs require use of induction on n in this and most of the following lemmas.)[1]

LEMMA 5. *If $a \succ b$, then there is a (positive integer) n such that aJ^nb.*

LEMMA 6. *If aJ^nb and aJ^nc, then $b \sim c$.*

LEMMA 7. *If aJ^mb and bJ^nc, then $aJ^{m+n}c$.*

LEMMA 8. *If aJ^mb and $aJ^{m+n}c$, then bJ^nc.*

LEMMA 9. *If $aJ^{m+n}b$, then there is a c in A such that aJ^mc.*

LEMMA 10. *If aJ^nb and cJ^nd, then $ab \sim cd$.*

LEMMA 11. *If $ab \sim cd$ then either there is some n such that aJ^nb and cJ^nd, or there is some n such that bJ^na and dJ^nc, or $a \sim b$ and $c \sim d$.*

We turn now to a sketch of the proof of Theorem 2. Let c^* be the first element of A with respect to the ordering $\succ$. Define the numerical function φ on A as follows for every a in A:

$$\varphi(a) = \begin{cases} 1 \textit{ if } a \sim c^*, \\ -n+1 \textit{ if } c^* J^n a. \end{cases}$$

Then using the elementary lemmas we may prove:

(i) $\varphi(a) > \varphi(b)$ if and only if $a \succ b$;

(ii) $\varphi(a) - \varphi(b) \succeq \varphi(d) - \varphi(e)$ if and only if $ab \succeq de$.

[1] In this and subsequent lemmas, as well as in the proof of Theorem 2 and later theorems, the concept of the nth power of the binary relation J is repeatedly used. This concept is defined recursively:

aJ^1b if and only if aJb,

aJ^nb if and only if there is a c such that $aJ^{n-1}c$ and cJb.

To prove that the function φ is unique up to a linear transformation, we define for every a in A two functions h_1 and h_2:

$$h_1(a) = \frac{\varphi_1(a) - \varphi_1(c^*)}{\varphi_1(c^*) - \varphi_1(c^{**})'}$$

$$h_2(a) = \frac{\varphi_2(a) - \varphi_2(c^*)}{\varphi_2(c^*) - \varphi_2(c^{**})'}$$

where φ_1 and φ_2 are two functions satisfying the representation construction and c^* is the first element of A under the ordering $\succeq$ and c^{**} the second element. We can easily show that h_1 is a linear transformation of φ_1 and h_2 is a linear transformation of φ_2 and also that h_1 is identical to h_2. It is then easy to prove that φ_1 is a linear transformation of φ_2, that is, there are numbers α, β with $\alpha > 0$ such that for every a in A

$$\varphi_1(a) = \alpha\,\varphi_2(a) + \beta.$$

Proof of Theorem 3. We begin with the proof of two lemmas. The first corresponds to Lemma 10 in the proof of Theorem 2 and the second to Lemma 11. It should be noted that the lemmas of Theorem 2 which are just about the relations $\succ$ and J also apply here.

LEMMA 1. *If aJ^nb and bJ^nc then $B(abc)$.*

Proof. We proceed by induction. For $n = 1$, we have Axiom 5. Suppose now that our inductive hypothesis holds and we have

(1) $aJ^{n+1}b$ *and* $bJ^{n+1}c$.

Then we know at once from properties of J that there are elements a' and c' in A such that

(2) aJa' *and* $a'J^nb$,

(3) bJ^nc' *and* $c'Jc$.

Whence by inductive hypothesis from (2) and (3)

(4) $B(a'bc')$,

and then from (2) and (3) again, as well as (4) and Axiom 6, we infer

$$B(abc)$$

as desired.

LEMMA 2. *If $B(abc)$ then there is an n such that aJ^nb and bJ^nc.*

Proof. From the hypothesis of the theorem and Axiom 4 we have

$$a \succ b \text{ and } b \succ c,$$

whence from familiar properties of J, there are m, n such that

$$aJ^mb \text{ and } bJ^nc.$$

Suppose now $m \neq n$; for definiteness and without loss of generality we may suppose that $m < n$. Then there is d such that

$$bJ^md,$$

whence by Lemma 1

$$B(abd),$$

but by hypothesis $B(abc)$, whence by Axiom 2

$$c \sim d.$$

But then we have

$$bJ^mc \textit{ and } bJ^nc,$$

which is impossible, and so we conclude $m = n$, as desired.

Given Lemmas 1 and 2, the proof of the existence of a function φ such that

(i) $\varphi(a) > \varphi(b)$ if and only if $a \succ b$

and

(ii) $\varphi(b) = \frac{1}{2}(\varphi(a) + \varphi(c))$ and $\varphi(a) > \varphi(b) > \varphi(c)$ if and only if $B(a, b, c)$

is similar to the proof of the corresponding part of Theorem 2 and need not be developed in detail.

For the proof of the uniqueness of φ up to a linear transformation, as in the case of the proof of Theorem 2, we assume we have two functions φ_1 and φ_2 both satisfying (i) and (ii). We then define h_1 and h_2, just as in that proof. By the very form of the definition it is clear that h_1 is a linear transformation of φ_1, and h_2 a linear transformation of φ_2. We complete the proof by an inductive argument to show that $h_1 = h_2$ (whence φ_2 is a linear transformation of φ_1).

The induction is with respect to the elements of A ordered by $\succ$, with c^* the first element.

Now by definition

$$h_1(c^*) = h_2(c^*) = 0.$$

Suppose now that for a_m, with $m \leq n$,

$$h_1(a_m) = h_2(a_m).$$

We prove that

$$h_1(a_{n+1}) = h_2(a_{n+1}).$$

Now we know at once that $a_{n-1}Ja_n$ and a_nJa_{n+1}, whence by virtue of Axiom 5

$$B(a_{n-1}, a_n, a_{n+1}),$$

and therefore by hypothesis

$$2\varphi_i(a_n) = \varphi_i(a_{n-1}) - \varphi_i(a_{n+1}),$$

whence

$$\varphi_i(a_{n+1}) = 2\varphi_i(a_n) - \varphi_i(a_{n-1}),$$

for $i = 1,2$. Now since h_i is a linear transformation of φ_i, it follows that we also have

$$h_i(a_{n+1}) = 2h_i(a_n) - h_i(a_{n-1}),$$

but by inductive hypothesis the right-hand side of this last equation is the same for h_1 and h_2, and so we conclude that $h_1(a_{n+1}) = h_2(a_{n+1})$.

Proof of Theorem 4. First of all, on the basis of Axioms 1–3 of Definition 9 the following elementary lemmas about the ordering induced on the two components A_1 and A_2 are easily proved.

LEMMA 1. *The relation $\sim$ on A_i, for i= 1, 2, is an equivalence relation, i.e., it is reflexive, symmetric, and transitive.*

LEMMA 2. *The relation $\succ$ on A_i, for $i = 1, 2$, is asymmetric and transitive.*

LEMMA 3. *For a and b in A_1 exactly one of the following is true: $a \sim b$, $a \succ b$, $b \succ a$. For p and q in A_2, exactly one of the following is true: $p \sim q$, $p \succ q$, $q \succ p$.*

We next prove the two lemmas mentioned earlier in the discussion of the axioms of Definition 9.

LEMMA 4. *If* $(a,p) \sim (b,q)$ *and* $a \succ b$ *then* $q \succ p$.

Proof. Suppose it is not the case that $q \succ p$. Then by Lemma 3 either $p \sim q$ or $p \succ q$. If $p \sim q$, then $(a,p) \sim (a,q)$, whence by transitivity and the hypothesis of the lemma, $(b,q) \sim (a,q)$, and thus $b \sim a$, which contradicts Lemma 3 and the hypothesis that $a \succ b$. On the other hand, a contradiction also follows from the supposition of the other alternative, i.e., $p \succ q$. For we have $(a,p) \succ (a,q)$, whence by familiar properties of weak orderings and the hypothesis of the lemma, $(b,q) \succ (a,q)$ and thus $b \succ a$, which again contradicts Lemma 3 and the hypothesis that $a \succ b$. Thus, we conclude that from the hypothesis of the lemma it follows that $q \succ p$, as desired.

LEMMA 5. *If* $(a,p) \sim (b,q)$ *and* $p \succ q$ *then* $b \succ a$.

Proof. Identical in structure to that for Lemma 4.

We turn next to the proof of Theorem 4. The proof closely resembles that of Theorem 2. Let c^* be the first element of A_1 with respect to the ordering $\succeq$ on A_1, and let r^* be the first element of A_2 with respect to the ordering $\succeq$ on A_2. Define, then, the numerical functions φ_1 and φ_2 on A_1 and A_2 as follows (for a in A_1 and p in A_2):

$$\varphi_1(a) = \begin{cases} 1 \text{ if } a \sim c^*, \\ -n+1 \text{ if } c^* J^n a, \end{cases}$$

$$\varphi_2(p) = \begin{cases} 1 \text{ if } pIr^*, \\ -n+1 \text{ if } r^* J^n p. \end{cases}$$

As in the case of the proof of Theorem 2, it is easy to show:

$$\varphi_1(a) > \varphi_1(b) \text{ if and only if } a \succ b,$$

$$\varphi_2(p) > \varphi_2(q) \text{ if and only if } p \succ q.$$

Moreover, Lemmas 1–9 proved in preparation for the proof of Theorem 2 also hold in the present setting, for they just depend on the binary relations on the components. Of course, for each of these lemmas, there is, strictly speaking, now a pair of lemmas, one for the ordering on each component.

Corresponding to Lemma 10 of this earlier list, we can now prove by the same inductive argument, using Axiom 4 of Definition 9:

(i) if $aJ^n b$ and $pJ^n q$ then $(a,q) \sim (b,p)$.

Second, we can prove the elementary fact:

(ii) if $(a,q) \sim (b,p)$ then either (a) there is some n such that $aJ^n b$ and $pJ^n q$, or (b) there is some n such that $bJ^n a$ and $qJ^n p$, or (c) $a \sim b$ and $p \sim q$.

From (i) and (ii) we prove then the fundamental result that $\varphi_1(a) + \varphi_2(q) = \varphi_1(b) + \varphi_2(p)$ if and only if $(a,q) \sim (b,p)$, which completes the first part of the proof of Theorem 4.

To prove the uniqueness results on φ_1 and φ_2, we may proceed as in the case of Theorem 2. We define four functions:

$$g(a) = \frac{\varphi_1(a) - \varphi_1(c^*)}{\varphi_1(c^*) - \varphi_1(c^{**})'} \qquad g'(a) = \frac{\varphi_1'(a) - \varphi_1'(c^*)}{\varphi_1'(c^*) - \varphi_1'(c^{**})'}$$

$$h(p) = \frac{\varphi_2(p) - \varphi_2(r^*)}{\varphi_2(r^*) - \varphi_2(r^{**})'} \qquad h'(p) = \frac{\varphi_2'(p) - \varphi_2'(r^*)}{\varphi_2'(r^*) - \varphi_2'(r^{**})'},$$

where c^* is the first element of A, under the ordering $\succeq$ on A_1, c^{**} is the second element, r^* is the first element of A_2, and r^{**} the second. It is, as before, obvious that g is a linear transformation of φ_1, g' a linear transformation of φ_1', h a linear transformation of φ_2, and h' a linear transformation of φ_2'. Secondly, we can show that $g = g'$ and $h = h'$ by an inductive argument similar to that used in the proof of Theorem 3. So we obtain that there are numbers α, α', β and γ with $\alpha, \alpha' > 0$ such that for every a in A_1 and every p in A_2

(iii) $\varphi_1'(a) = \alpha\varphi_1(a) + \beta$ and $\varphi_2'(p) = \alpha'\varphi_2(p) + \gamma$.

It remains to show that $\alpha = \alpha'$ when A_1 and A_2 each have at least two elements not equivalent in order. Without loss of generality we may take $a \succ b$ and $p \succ q$. Then we have, from $(a,q) \sim (b,p)$,

$$\varphi_1'(a) - \varphi_1'(b) = \varphi_2'(p) - \varphi_2'(q),$$

and thus by (iii)

$$\frac{\alpha\varphi_1(a) - \alpha\varphi_1(b)}{\alpha'\varphi_2(p) - \alpha'\varphi_2(q)} = 1,$$

and so

$$\frac{\alpha}{\alpha'}\left(\frac{\varphi_1(a) - \varphi_1(b)}{\varphi_2(p) - \varphi_2(q)}\right) = 1;$$

but by hypothesis

$$\varphi_1(a) - \varphi_1(b) = \varphi_2(p) - \varphi_2(q),$$

whence

$$\frac{\alpha}{\alpha'} = 1;$$

i.e.,

$$\alpha = \alpha',$$

which completes the proof.

14

THE MEASUREMENT OF BELIEF

1. INTRODUCTION

Almost everyone who has thought about the problems of measuring beliefs in the tradition of subjective probability or Bayesian statistical procedures concedes some uneasiness with the problem of always asking for the next decimal of accuracy in the prior estimation of a probability or of asking for the parameter of a distribution that determines the probabilities of events. On the other hand, the formal theories that have been developed for rational decision-making under uncertainty by Ramsey (1951), de Finetti (1931, 1937), Koopman (1940a, b), Savage (1954) and subsequent authors have almost uniformly tended to yield a result that guarantees a unique probability distribution on states of nature or whatever other collection of entities is used for the expression of prior beliefs.

In the next section I examine some of these standard theories and address the question of how we can best criticize the claims they make. Among other points, I consider the claim that the idealizations expressed in the axioms can be regarded as theories of pure rationality.

In the third section I examine two constructive possibilities that yield inexact measurements of belief. Because the issues are almost entirely conceptual and not technical at the present stage of investigation, I confine myself to comparing some elementary axiom systems and raise the

*Reprinted from *Journal of the Royal Statistical Society* (Series B), **36** (1974), 160-191. I am indebted to Mario Zanotti for a number of useful comments on the ideas developed here, especially those dealing with upper and lower probability.

question of their suitability as a basis for empirical investigation. The first system is relatively trivial, but is designed to make a certain conceptual point. The second system is considerably more interesting and presents, I think, a useful approach to inexact measurement of subjective probability, with a representation theorem formulated in terms of upper and lower probabilities.

In the final section I compare the measurement of belief to the classical theory of measurement embodied in Euclidean geometry and challenge the view that idealizations of exact measurement are as useful and harmless in the case of the theory of beliefs as they are in the case of geometry. I also briefly compare the situation with that which exists in quantum mechanics and meteorology and argue for the conclusion that the inexact results of these sciences are a more appropriate model than that of geometry. More importantly, I try to state in this section some unfinished ideas about processes for constructing beliefs.

I do not have as much to say about empirical matters in this article as I would like. A common view I share is that the conceptual and formal analysis of belief structures has currently far outstripped the empirical study of beliefs, and probably what is needed most at the present time are several relentless programs of empirical investigation guided and motivated by the insights afforded from various formal concepts and theories that are mathematically now well understood.

2. WEAKNESSES OF THE STANDARD THEORIES

Because the standard theories mentioned earlier reach essentially the same formal results, namely, the existence of a unique probability distribution on states of nature, criticisms of one will pretty much apply to criticisms of the lot. For this reason, it may pay to concentrate on Savage's (1954) axioms, because of their familiarity to a wide audience and because they have been much discussed in the literature. I emphasize, however, that what I have to say about Savage's axioms will apply essentially without change to other standard theories.

Because Savage's axioms are rather complicated from a formal standpoint, I shall not state them explicitly here, but shall try to describe their intuitive content. The axioms are about preference among decisions, where decisions are mappings or functions from the set of states of nature to the set of consequences. To illustrate these ideas, let me use an example I have used before (Suppes, 1956).

A certain independent distributor of bread must place his order for a given day by ten o'clock of the preceding evening. His sales to inde-

	d_1 buy 700 loaves	d_2 buy 800 loaves	d_3 buy 900 loaves
s_1–rain	$21.00	$19.00	$17.00
s_2–no rain	$21.00	$24.00	$26.50

Table 14.1.

pendent grocers are affected by whether or not it is raining at the time of delivery, for if it is raining, the grocers tend to buy less on the accumulated evidence that they have fewer customers. On a rainy day the maximum the distributor can sell is 700 loaves; on such a day he makes less money if he has ordered more than 700 loaves. On the other hand, when the weather is fair, he can sell about 900 loaves. If the simplifying assumption is made that the consequences to him of a given decision with a given state of nature (s_1–rain or s_2–no rain) may be summarized simply in terms of his net profits, the situation facing him is represented in Table 1. The distributor's problem is to make a decision.

Clearly, if he knows for certain that it is going to rain, he should make decision d_1, and if he knows for certain that it is not going to rain, he should make decision d_3. The point of Savage's theory, expanded to more general and more complex situations, is to place axioms on choices or preferences among the decisions in such a way that anyone who satisfies the axioms will be maximizing expected utility. This means that the way in which he satisfies the axioms will generate a subjective probability distribution about his beliefs concerning the true state of nature and a utility function on the set of consequences such that the expectation of a given decision is defined in a straightforward way with respect to the subjective probability distribution on states of nature and the utility function on the set of consequences. As one would expect, Savage demands, in fact in his first axiom, that the preference among decisions be transitive and that given any two decisions one is at least weakly preferred to the other. Axiom 2 extends this ordering assumption to having the same property hold when the domain of definition of decisions is restricted to a given set of states of nature; for example, the decision-maker might know that the true state of nature lies in some subset of the whole set. Axiom 3 asserts that knowledge of an event cannot change preferences among consequences, where preferences among consequences are defined in terms of preferences among decisions. Axiom 4 requires that given any two sets of states of nature, that is, any two events, one is at least as probable as the other, that is, qualitative probability among events is strongly connected. Axiom 5 excludes the trivial case in which all consequences are equivalent in utility and, thus, every decision is equivalent to every other. Axiom 6 says essentially that if event A is less probable than event B (A and B are subsets of the same set of states of nature), then there is a partition of the states of nature such that the union of each element of the partition with A is less probable than B. As is well known, this axiom of Savage's is closely related to the axiom of de Finetti and Koopman, which requires the existence of a partition of the states of nature into arbitrarily many events that are equivalent in probability. Finally, his last axiom, Axiom 7, is a formulation of the sure-thing principle.

My first major claim is that some of Savage's axioms do not in any direct sense represent axioms of rationality that should be satisfied by any ideally rational person but, rather, they represent structural assumptions about the environment that may or may not be satisfied in given applications.

Many years ago, at the time of the Third Berkeley Symposium (1955), I introduced the distinction between structure axioms and rationality axioms in the theory of decision-making (Suppes, 1956). Intuitively, a structure axiom as opposed to a rationality axiom is existential in character. In the case of Savage's seven postulates, two (5 and 6) are structure axioms, because they are existential in character.

Savage defended his strong Axiom 6 by holding it applicable if there is a coin that a decision-maker believes is fair for any finite sequence of flips. There are however, several objections to this argument. First of all, if it is taken seriously then one ought to redo the entire foundations and simply build it around Bernoulli sequences with $p = 0 \cdot 5$ and get arbitrarily close approximations to the probability of any desired event. (See the second system of axioms in the next section.) More importantly, without radical changes in human thinking, it is simply not natural on the part of human beings to think of finite sequences of flips of a coin in evaluating likelihoods or probabilities, qualitative or quantitative, of significant events with which they are concerned.

Consider the case of a patient's deciding whether to follow a surgeon's advice to have major surgery. The surgeon, let us suppose, has evaluated the pros and cons of the operation, and the patient is now faced with the critical decision of whether to take the risk of major surgery with at least a positive probability of death, or whether to take the risk of having no surgery and suffering the consequences of the continuing disease. I find it very unlikely and psychologically very unrealistic to believe that thinking about finite sequences of flips of a fair coin will be of any help in making a rational decision on the part of the patient.

On the other hand, other axioms like those on the ordering of preferences or qualitative probability seem reasonable in this framework and are not difficult to accept. But the important point is this. In a case in which uncertainty has a central role, in practice, decisions are made without any attempt to reach the state of having a quantitative probability estimate of the alternatives or, if you like, a computed expected utility.

It is, in fact, my conviction that we usually deal with restricted situations in which the set of decisions open to us is small and in which the events that we consider relevant are small in number. The kind of enlarged decision framework provided by standard theories is precisely the source of the uneasiness alluded to in the first sentence of the introduction. Intuitively we all move away from the idea of estimating probabilities with arbitrary refinement. We move away as well from the introduction of an elaborate mechanism of randomization in order to have a sufficiently large decision space. Indeed, given the Bayesian attitude towards randomization, there is an air of paradox about the introduction *à la* Savage of finite sequences of tosses of a fair coin.

Another way of putting the matter, it seems to me, is that there is a strong intuitive feeling that a decision-maker is not irrational simply because a wide range of decision possibilities or events is not available to him. It is not a part of rationality to require that the decision-maker enlarge his decision space, for example, by adding a coin that may be flipped any finite number of times. I feel that the intrinsic theory of rationality should be prepared to deal with a given set of states of nature and a given set of decision functions, and it is the responsibility of the formal theory of belief or decision to provide a theory of how to deal with these restricted situations without introducing strong structural assumptions.

A technical way of phrasing what I am saying about axioms of pure rationality is the following. For the moment, to keep the technical apparatus simple, let us restrict ourselves to a basic set S of states of nature and a binary ordering relation of qualitative probability on subsets of S, with the usual Boolean operations of union, intersection and complementation having their intuitive meaning in terms of events. I then say that an axiom about such structures is an axiom of pure rationality only if it is closed under submodels. Technically, closure under submodels means that if the axiom is satisfied for a pair $\langle S, \succeq \rangle$ then it is satisfied for any nonempty subset of S with the binary relation $\succeq$ restricted to the power set of the subset, i.e., restricted to the set of all subsets of the given subset. (Of course, the operations of union, intersection and complementation are closed in the power set of the subset.) Using this technical definition, we can easily see that of Savage's seven axioms, five of them satisfy this restriction, and the two already mentioned as structure axioms do not.

Let me try to make somewhat more explicit the intuition which is behind the requirement that axioms of pure rationality should satisfy the condition of closure under submodels. One kind of application of the condition is close to the axiom on the independence of irrelevant alternatives in the theory of choice. This axiom says that if we express a preference among candidates for office, for example, and if one candidate is removed from the list due to death or for other reasons, then our ordering of preferences among the remaining candidates should be unchanged. This axiom satisfies closure under submodels. The core idea is that existential requirements that reach out and make special requirements on the environment do not represent demands of pure rationality but rather structural demands on the environment, and such existential demands are ruled out by the condition of closure under submodels.

A different, but closely related, way of defining axioms of pure rationality is that such an axiom must be a logical consequence of the existence of the intended numerical representation. This criterion, which I shall call the *criterion of representational consequence*, can be taken as both necessary and sufficient, whereas the criterion of closure under submodels is obviously not sufficient. On the other hand, the extrinsic character of the criterion of representational consequence can be regarded as unsatisfactory. It is useful for identifying axioms that are not necessary for the intended representation and thus smuggle in some unwanted arbitrary structural assumption. As should be clear, Savage's Axioms 5 and 6 do such smuggling.

I am quite willing to grant the point that axioms of rationality of a more restricted kind could be considered. One could argue that we need special axioms of rationality for special situations, and that we should embark on a taxonomy of situations providing appropriate axioms for each of the major classes of the taxonomy. In the present primitive state of analysis, however, it seems desirable to begin with a sharp distinction between rationality and structure axioms and to have the concept of pure rationality universal in character.

Returning now to my criticisms of Savage's theory, it is easy to give finite or infinite models of Savage's five axioms of rationality for which there exists no numerical representation in terms of utility and subjective probability. In the language I am using here, Savage's axioms of pure rationality are insufficient for establishing the existence of representing numerical utility and subjective probability functions.

Moreover, we may show that no finite list of additional elementary axioms of a universal character will be sufficient to guarantee the existence of appropriate numerical functions. By *elementary axioms* I mean axioms that can be expressed within first-order logic. First-order logic essentially consists of the conceptual apparatus of sentential connectives, one level of variables and quantifiers for these variables, together with non-logical predicates, operation symbols and individual constants. Thus, for example, the standard axioms for groups or for ordered algebraic fields are elementary, but the least upper-bound axiom for the field of real numbers is not. It is possible to formulate Savage's Axiom 5 in an elementary way, but not his Axiom 6.

In the case of infinite models, the insufficiency of elementary axioms, without restriction to their being of a universal character, follows from the upward Lowenheim-Skolem-Tarski theorem, plus some weak general assumptions. This theorem asserts that if a set of elementary axioms has an infinite model (i.e., a model whose domain is an infinite set, as is the case for Savage's theory), then it has a model of every infinite cardinality. Under quite general assumptions, e.g., on the ordering relation of preference or greater subjective probability, it is impossible to map the models of high infinite cardinality into the real numbers, and thus no numerical representation exists.

In the case of finite models, the methods of Scott and Suppes (1958) apply to show that no finite set of universal elementary axioms will suffice. The system consisting of Savage's five axioms of pure rationality has finite models, but by the methods indicated we can show there is no finite elementary extension by means of universal axioms of rationality that will be strong enough to lead to the standard numerical representation. (The essential idea of Scott and Suppes' work is to show that if for every positive integer n there is a finite model M such that every submodel of n elements satisfies the theory in question, but the model M does not, then the theory is not axiomatizable by a finite list of elementary axioms that are universal in form.)

The results I have outlined indicate the nature of some of the general restrictions that obtain in the hope of finding elementary axioms of pure rationality sufficient to lead to an appropriate numerical representation of the decision situation.

On the other hand, in the case of finite models, necessary and sufficient conditions can be given, and using the criterion of closure under submodels as a criterion of pure rationality, we then have formally adequate axioms of pure rationality in the finite case, even if the conditions are not fixed in number, but are represented by a potentially infinite schema.

The simplest and most elegant version of such axioms is probably that given by Scott (1964) for the de Finetti framework of qualitative subjective probability in which decisions and consequences are not explicitly considered. (His axioms improve on the earlier ones given by Kraft et al., 1959.) Because I want to comment on their character from the standpoint developed in this paper, Scott's axioms are embodied in the following definition, in which the notation A^c is used for the characteristic function of a set A, and $\emptyset$ for the empty set.

DEFINITION 1. *Let X be a non-empty finite set and $\succeq$ a binary relation on the set of all subsets of X. Then a structure $\langle X, \succeq \rangle$ is* a (finite) qualitative belief structure *if and only if for all subsets A and B of X*

Axiom 1. $A \succeq B$ *or* $B \succeq A$;

Axiom 2. $A \succeq \emptyset$;

Axiom 3. $X \succ \emptyset$;

Axiom 4. *For all subsets $A_0, \ldots, A_n, B_0, \ldots, B_n$ of X, if $A_i \succeq B_i$ for $0 \leq i < n$, and for all x in X*

$$A_0^c(x) + \ldots + A_n^c(x) = B_0^c(x) + \ldots + B_n^c(x),$$

then $B_n \succeq A_n$.

Axiom 4 only requires that any element of X, that is, any atomic event, belong to exactly the same number of A_i and B_i, for $0 \leq i \leq n$. To illustrate the force of Scott's Axiom 4, we may see how it implies transitivity. First, necessarily for any three characteristic functions

$$A^c + B^c + C^c = B^c + C^c + A^c;$$

that is, for all elements x of X

$$A^c(x) + B^c(x) + C^c(x) = B^c(x) + C^c(x) + A^c(x).$$

By hypothesis, $A \succeq B$ and $B \succeq C$, whence by virtue of Axiom 4,

$$C \preceq A,$$

and thus, by definition $A \succeq C$, as desired. Scott proves that for any finite structure $\mathcal{X} = \langle X, \succeq \rangle$ satisfying the axioms of Definition 1 there is a probability measure P such that for A and B subsets of X

$$A \succeq B \text{ if and only if } P(A) \geq P(B).$$

A first point to note is that the probability measure P is not unique, nor apparently can its uniqueness up to a given set of transformations be characterized in an interesting way, a situation that is true for many finite geometries when the set of transformations is as general as possible consistent with the finite number of relationships expressed.

The more profound difficulty with Scott's axioms as a theory of belief is the combinatorial explosion that occurs in verifying the axioms when the number of events is large. To check connectedness, for example, we need only consider pairs of events, and to check transitivity, only triples of events. But, it is fundamental for the kind of axiom schema (Scott's Axiom 4) required to express necessary and sufficient conditions in the finite case that n-tuples of events of arbitrary n must be studied as the number of events increases. As a possible empirical theory of belief, or as a rational one, this seems impractical, and even for fairly small experiments, the effort to determine whether there is a representing probability measure requires the use of a moderate-sized computer facility. Certainly the experiments do not themselves check all the possible n-tuples of comparison. Again, I will not enter into detailed computations, but in conducting some unpublished experiments on measuring beliefs some years ago, already I found that in considering a space with ten atoms, a small number for complex matters, the combinatorial explosion of possible comparisons of pairs of events (not necessarily atomic) was impressive. (Talk about atoms is just another way of talking about the points in a sample space.) If we deal with 30 or 40 or 50 atoms, the numbers are out of hand, even when we take maximal advantage of relationships implied by the axioms.

3. INEXACT MEASUREMENT

In thinking about these problems once again, I asked myself what are the simplest axioms that would minimize the number of comparisons needed, and that would still yield some results on the underlying measure if it is there. You may find the following axioms amusing. Although I do not propose them as a serious set to be used in extensive studies of actual beliefs, I do advance them as one modest conceptual model of how far we can go in simplifying the comparisons we ask for, and yet obtain some kind of results different from those of simple order if the axioms are satisfied.

The intuitive idea of the restricted system is to have five classes of events: Those that are certain (C), those that are more likely than not (M), those that are less likely than not (L), those that are as likely as not (E) and those that are impossible (I). However, only two of these five classes of events need be taken as primitive. For example, taking the class of certain events and the class of events that are more likely than not as primitive, we can define the other three in the following manner, where if A is an event, then *not* A is of course the event that occurs if A does not, i.e., the complement of A: A *is impossible* if and only if *not* A is certain; A *is less likely than not* if and only if *not* A is more likely than not; A *is as likely as not* if and only if A is neither certain, impossible, more likely than not, nor less likely than not.

Let X be a non-empty set and let events be subsets of X. Then the axioms of what I shall call *weak qualitative probability structures* are the following:

Axiom 1. X is certain.

Axiom 2. If A implies B and A is certain, then B is certain.

Axiom 3. If A implies B and A is more likely than not, then B is more likely than not.

Axiom 4. If A implies B but B does not imply A and A is as likely as not, then B is more likely than not.

Axiom 5. If A is certain, then *not* A is impossible.

Axiom 6. If A is more likely than not, then *not* A is less likely than not.

(A completely formal version of these axioms can easily be given.)

From the axioms we can easily prove the following sorts of elementary theorems: If A implies B and B is less likely than not, then A is less likely than not; if A is as likely as not, then *not* A is as likely as not; if A is as likely as not, B is as likely as not, and A and B are mutually exclusive, then the disjunction A or B is certain. (The proof of the last assertion uses Axiom 4.)

In many cases the situation described by these axioms is about the appropriate degree of crudeness of what a person knows about his beliefs. Even in the present framework we can add axioms that will force the situation to be much tighter. These axioms are of course structural axioms and in general will not be satisfied in a given situation. For example, we can require that every atom be less likely than not, but still not impossible, and also that if an event is less likely than not, then there is some second event such that the disjunction of the two is as likely as not. When these structural assumptions are added, we can show that a system with three atoms is impossible, and a system of four atoms requires that they be equally probable. The three-atom case is easy to see. By way of contradiction, let x, y and z be the numerical probabilities of the three atoms. By hypothesis $x < \frac{1}{2}$, and thus also by hypothesis either $x+y = \frac{1}{2}$ or $x + z = \frac{1}{2}$, but in the first case then $z = \frac{1}{2}$, contrary to assumption, and in the second case, $y = \frac{1}{2}$, also contrary to assumption. I shall not explore the situation in more detail, because it seems to me that these particular structural axioms are not especially interesting. I merely state them as an indication of the kind of results we can get by some relatively innocent-appearing structural assumptions. Notice that even with the structural atoms, we are not able to prove that there is an ordering of events in terms of less probable and more probable.

For weak qualitative probability structures, we can prove a representation theorem.

THEOREM 1. *If X is finite or countable, and $\langle X, C, M\rangle$ is a weak qualitative probability structure, then there is a probability measure defined on the power set of X such that*

(i) $P(A) = 1$ *if and only if A is certain,*

(ii) $P(A) > \frac{1}{2}$ *if and only if A is more likely than not.*

From the definitions given above it follows that $P(A) < \frac{1}{2}$ if and only if A is less likely than not, $P(A) = \frac{1}{2}$ if and only if A is as likely as not and $P(A) = 0$ if and only if A is impossible.

As a final system of axioms, I want to introduce purely in terms of belief or subjective probability what I consider the appropriate finitistic analogue of Savage's axioms. These constitute an extension of de Finetti's qualitative conditions and lead to simple approximate measurement of belief in arbitrary events. The axioms require something that I partly criticized earlier, namely, the existence of some standard set of events whose probability is known exactly. They would, for example, be satisfied by flipping a fair coin n times for some fixed n. They do not require that n be indefinitely large and therefore n may be looked upon as somewhat more realistic. I give the axioms here in spite of my feeling that, from the standpoint of a serious decision like that on surgery mentioned earlier, they may be unsatisfactory.

They do provide a combination of de Finetti's ideas and a finite version of the standard structural axiom on infinite partitions.

. The concept of upper and lower probabilities seems to be rather recent in the literature, but it is obviously closely related to the classical concepts of inner and outer measure, which were introduced by Caratheodory and others at the end of the nineteenth century and the beginning of this century. Koopman (1940b) explicitly introduces lower and upper probabilities but does nothing with them from a conceptual standpoint. He uses them as a technical device, as in the case of upper and lower measures in mathematical analysis, to define probabilities. The first explicit conceptual discussions seem to be quite recent (Smith, 1961; Good, 1962). Smith especially enters into many of the important conceptual considerations, and Good states a number of the quantitative properties it seems natural to impose on upper and lower probabilities. Applications to problems of statistical inference are to be found in Dempster (1967). However, so far as I know, a simple axiomatic treatment starting from purely qualitative axioms does not yet exist in the literature, and the axioms given below represent such an effort. It is apparent that they are not the most general axioms possible, but they do provide a simple and hopefully rather elegant qualitative base.

From a formal standpoint, the basic structures to which the axioms apply are quadruples $\langle X, \mathcal{F}, \mathcal{L}, \succeq\rangle$, where X is a non-empty set, $\mathcal{F}$ is an algebra of subsets of X, that is, $\mathcal{F}$ is a non-empty family of subsets of X and is closed under union and complementation, $\mathcal{L}$ is a similar algebra of sets, intuitively the events that are used for standard measurements, and I shall refer to the events in $\mathcal{L}$ as *standard* events S, T, etc. The relation $\succeq$ is the familiar ordering relation on $\mathcal{F}$. I use familiar abbreviations for equivalence and strict ordering in terms of the weak ordering relation. (A weak ordering is transitive and strongly connected, i.e., for any events A and B, either $A \succeq B$ or $B \succeq A$.)

DEFINITION 2. *A structure $\mathcal{X} = \langle X, \mathcal{F}, \mathcal{S}, \succeq\rangle$ is a* finite approximate measurement structure for beliefs *if and only if X is a non-empty set, $\mathcal{F}$ and $\mathcal{S}$ are algebras of sets on X, and the following axioms are satisfied for every A, B and C in $\mathcal{F}$ and every S and T in $\mathcal{S}$:*

Axiom 1. *The relation $\succeq$ is a weak ordering of $\mathcal{F}$;*

Axiom 2. *If $A \cap C = \emptyset$ and $B \cap C = \emptyset$ then $A \succeq B$ if and only if $A \cup C \succeq B \cup C$;*

Axiom 3. $A \succeq \emptyset$;

Axiom 4. $X \succ \emptyset$;

Axiom 5. *$\mathcal{S}$ is a finite subset of $\mathcal{F}$;*

Axiom 6. *If $S \neq \emptyset$ then $S \succ \emptyset$;*

Axiom 7. *If $S \succeq T$ then there is a V in $\mathcal{S}$ such that $S \approx T \cup V$.*

In comparing Axioms 3 and 6, note that A is an arbitrary element of the general algebra $\mathcal{F}$, but event S (referred to in Axiom 6) is an arbitrary element of the subalgebra $\mathcal{S}$. Also in Axiom 7, S and T are standard events in the subalgebra $\mathcal{S}$, not arbitrary events in the general algebra. Axioms 1–4 are just the familiar de Finetti axioms without any change. Because all the standard events (finite in number) are also events (Axiom 5), Axioms 1–4 hold for standard events as well as arbitrary events. Axiom 6 guarantees that every minimal element of the subalgebra $\mathcal{S}$ has positive qualitative probability. Technically a minimal element of $\mathcal{S}$ is any event A in $\mathcal{S}$ such that $A \neq \emptyset$, and it is not the case that there is a non-empty B in $\mathcal{S}$ such that B is a proper subset of A. A *minimal open interval* (S, S') of $\mathcal{S}$ is such that $S \prec S'$ and $S' - S$ is equivalent to a minimal element of $\mathcal{S}$. Axiom 7 is the main structural axiom, which holds only for the subalgebra and not for the general algebra; it formulates an extremely simple solvability condition for standard events. It was stated in this form in Suppes (1969b, p. 6) but in this earlier case for the general algebra $\mathcal{F}$.

In stating the representation and uniqueness theorem for structures satisfying Definition 3, in addition to an ordinary probability measure on the standard events, I shall use upper and lower probabilities to express the inexact measurement of arbitrary events. A good discussion of the quantitative properties one expects of such upper and lower probabilities is found in Good (1962). All of his properties are not needed here because he dealt with conditional probabilities. The following properties are fundamental, where $P_*(A)$ is the lower probability of an event A and $P^*(A)$ is the upper probability (for every A and B in $\mathcal{F}$):

I. $P_*(A) \geq 0$.

II. $P_*(X) = P^*(X) = 1$.

III. If $A \cap B = \emptyset$ then

$$P_*(A) + P_*(B) \leq P_*(A \cup B) \leq P_*(A) + P^*(B) \leq P^*(A \cup B) \\ \leq P^*(A) + P^*(B).$$

Condition (I) corresponds to Good's Axiom D2 and (III) to his Axiom D3.

For standard events $P(S) = P_*(S) = P^*(S)$. For an arbitrary event A not equivalent in qualitative probability to a standard event, I think of its "true" probability as lying in the open interval $(P_*(A), P^*(A))$.

Originally I included as a fourth property

$$P_*(A) + P^*(\neg A) = 1,$$

where $\neg A$ is the complement of A, but Mario Zanotti pointed out to me that this property follows from (II) and (III) by the following argument:

$$1 = P_*(X) = P_*(A \cup \neg A) \leq P_*(A) + P^*(\neg A) \leq P^*(A \cup \neg A) = P^*(X) = 1.$$

A stronger property possessed by some upper and lower measures is this:

IV. $P_*(A \cup B) + P_*(A \cap B) \geq P_*(A) + P_*(B)$.

Good mentions that he suspected that this principle is independent of the others he introduces. (He actually states the dual form in terms of upper probabilities.) After the proof of Theorem 2, I give a counterexample to show that (IV) does not hold for every qualitative structure satisfying Definition 3.

In the fourth part of Theorem 2, I define a certain relation and state it is a semiorder with an implication from the semiorder relation holding to an inequality for upper and lower probabilities. Semiorders have been fairly widely discussed in the literature as a generalization of simple orders first introduced by Duncan Luce. I use here the axioms given by Scott and Suppes (1958). A structure $\langle U, R\rangle$ where U is a non-empty set and R is a binary relation on U is a *semiorder* if and only if for all $x, y, z, w \in U$:

Axiom 1 . Not xRx;

Axiom 2. If xRy and zRw then either xRw or zRy;

Axiom 3. If xRy and yRz then either xRw or wRz.

THEOREM 2. *Let* $\mathcal{X} = \langle X, \mathcal{F}, \mathcal{S}, \succeq\rangle$ *be a finite approximate measurement structure for beliefs. Then*

(i) *there exists a probability measure* P *on* $\mathcal{S}$ *such that for any two standard events* S *and* T

$$S \succeq T \textit{ if and only if } P(S) \geq P(T),$$

(ii) *the measure* P *is unique and assigns the same positive probability to each minimal event of* $\mathcal{S}$,

(iii) *if we define* P_* *and* P^* *as follows:*

(a) *for any event* A *in* $\mathcal{F}$ *equivalent to some standard event* S,

$$P_*(A) = P^*(A) = P(S),$$

(b) *for any* A *in* $\mathcal{F}$ *not equivalent to some standard event* S, *but lying in the minimal open interval* (S, S') *for standard events* S *and* S'

$$P_*(A) = P(S) \text{ and } P^*(A) = P(S'),$$

then P_* *and* P^* *satisfy conditions (I)–(III) for upper and lower probabilities on* $\mathcal{F}$, *and*

(c) *if* n *is the number of minimal elements in* $\mathcal{S}$ *then for every* A *in* $\mathcal{F}$

$$P^*(A) - P_*(A) \leq 1/n,$$

(iv) *if we define for A and B in $\mathcal{F}$*

$$A{*}\succ B \text{ if and only if } \exists S \text{ in } \mathcal{S} \text{ such that } A \succ S \succ B,$$

then $*\succ$ *is a semiorder on* $\mathcal{F}$, *if* $A{*}\succ B$ *then* $P_*(A) \geq P^*(B)$, *and if* $P_*(A) \geq P^*(B)$ *then* $A \succeq B$.

Proof. Parts (i) and (ii) follow from the proof given in Suppes (1969, pp. 7–8) once it is observed that the subalgebra $\mathcal{S}$ is isomorphic to a finite algebra $\mathcal{U}$ of sets with the minimal events of $\mathcal{S}$ corresponding to unit sets, i.e., atomic events of $\mathcal{U}$.

As to part (iii), conditions (I) and (II) for upper and lower probabilities are verified immediately. To verify condition (III) it will be sufficient to assume that neither A nor B is equivalent to a standard event, for if either is, the argument given here is simplified, and if both are, (III) follows at once from properties of the standard measure P. So we may assume that A is in a minimal interval (S, S') and B in a minimal interval (T, T'), i.e., $S \prec A \prec S'$ and $T \prec B \prec T'$. Since by hypothesis of (III), $A \cap B = \emptyset, T \preceq \neg S$ for if $T \succ \neg S$, we would have $A \cup B \succ S \cup \neg S$, which is impossible. Now it is easily checked that for standard events if $T \preceq \neg S$ then $\exists T^*$ in $\mathcal{S}$ such that $T^* \approx T$ and $T^* \subseteq \neg S$. So we have

$$P_*(A) + P_*(B) \leq P(S) + P(T^*) = P(S \cup T^*) \leq P_*(A \cup B),$$

with the last inequality following from $S \cup T^* < A \cup B$, which is itself a direct consequence of $S \prec A, T^* \prec B, A \cap B = \emptyset$ and Axiom 2. For the next step, if $\exists T^{**}$ in $\mathcal{S}$ such that $T^{**} \approx T'$ and $T^{**} \subseteq \neg S'$, then $A \cup B \prec S' \cup T^{**}$ and let $A \cup B$ be in the minimal closed interval $[V, V']$, i.e., $V \preceq A \cup B \preceq V'$. Then it is easy to show that $V \preceq S \cup T^{**}$, whence

$$P_*(A \cup B) = P(V) \leq P(S \cup T^{**}) = P(S) + P(T^{**}) = P_*(A) + P^*(B)$$

and since $S \cup T^* \prec A \cup B$, and $V \preceq S \cup T^{**}$, either $A \cup B \preceq S \cup T^{**}$ or $A \cup B \preceq S' \cup T^{**}$. In either case

$$\begin{aligned} P_*(A) + P^*(B) &= P(S \cup T^{**}) \leq P^*(A \cup B) \leq P(S' \cup T^{**}) \\ &= P(S') + P(T^{**}) = P^*(A) + P^*(B). \end{aligned}$$

On the other hand, if there were no T^{**} such that $T^{**} \approx T'$ and $T^{**} \subseteq \neg S'$, then $T' \succ \neg S'$, so that $S \cup T^* = S \cup \neg S$, and consequently $A \cup B \approx S \cup T^*$, so that $A \succeq S$ or $B \succeq T^*$ contrary to hypothesis, which completes the proof of (III).

Proof of (c) of part (iii) follows at once from (ii) and the earlier parts of (iii). Proof of (iv) is also straightforward and will be omitted.

I turn now to some remarks about Theorem 2. The implications stated in part (iv) cannot be strengthened to equivalence. It is easy to give counterexamples to each of the following four equivalences:

$A \mathrel{*\succ} B$ if and only if $P_*(A) \geq P^*(B)$;
$A \succeq B$ if and only if $P_*(A) \geq P^*(B)$;
$A \succ B$ if and only if $P_*(A) \geq P^*(B)$;
$A \succ B$ if and only if $P_*(A) > P^*(B)$.

A counterexample to the strong condition (IV) for upper and lower probabilities is the following. Let the outcomes of X be the four possible outcomes of two flips of a coin, the first without bias and the second with some unknown bias favouring heads. Explicitly, let $X = \{hh, ht, th, tt\}$. Then the standard events are $X, \emptyset, \{hh, ht\}$ and $\{th, tt\}$, with $P(\{hh, ht\}) = P(\{th, tt\}) = \frac{1}{2}$. Let $A = \{ht, hh\}$ and $B = \{hh, tt\}$. Then it is easy to see that $P_*(A) = P_*(B) = P_*(A \cup B) = \frac{1}{2}$, but $P_*(A \cap B) = 0$, and thus (IV) does not hold.

In my opening remarks I mentioned the embarrassing problem of being asked for the next decimal of a subjective probability. Without claiming to have met all such problems, the results embodied in Theorem 2 show that the axioms of Definition 3 provide a basis for a better answer. If there are n minimal standard events, then the probabilities of the 2^n standard events are known exactly as rational numbers of the form m/n, with $0 \leq m \leq n$, and further questions about precision are mistaken. The upper and lower probabilities of all other events are defined in terms of the probabilities of the 2^n standard events, and so the upper and lower probabilities are also known exactly as rational numbers of the same form m/n.

One can object to knowing the probabilities of standard events exactly, but this is to raise another problem that I also think can be dealt with in a way that improves on the axioms of Definition 3, but these additional matters will have to be pursued on another occasion.

Finally, I note explicitly that there is no need in Definition 3 to require that the sample space X be finite. The only essential requirement is that the set $\mathcal{S}$ of standard events be finite. The algebra $\mathcal{F}$ could even have a cardinality greater than that of the continuum and thus the order relation $\succeq$ on $\mathcal{F}$ might not be representable numerically, and yet the upper and lower probabilities for all events in $\mathcal{F}$ would exist and be defined as in the theorem.

4. COMPARISON WITH GEOMETRY

I mentioned at the beginning that I wanted to compare the measurement of belief with the kind of classical measurement characteristic of geometry. We are all familiar with what we expect of geometry, namely, that sufficient postulates are laid down to lead to a unique representation of the Euclidean plane or Euclidean space in terms of numerical Cartesian coordinates. The theory leads to exact results, and the uniqueness of the measurements, that is, the numbers assigned to points, is determined up to the group of rigid motions. This is a seductive ideal and is often taken as the ideal we should aim at in the case of the measurement of belief.

My point is to express skepticism that this is the correct ideal and to conjecture that the situation is more like the prototypical situation of quantum mechanics. Any time we measure a microscopic object by using macroscopic apparatus we disturb the state of the microscopic object and, according to the fundamental ideas of quantum mechanics, we cannot hope to improve the situation by using new methods of measurement that will lead to exact results of the classical sort for simultaneously measured conjugate variables. I do not mean to suggest that the exact theoretical ideas of quantum mechanics carry over in any way to the measurement of belief, but I think the general conceptual situation does. In fact, it seems to me that some of the recent empirical work of Tversky and his collaborators shows how sensitive the measurement of belief in the sense of subjective probability is to the particular method of measurement chosen. There is a general way of talking about this situation that is suggestive of a line of investigation, in terms of the theory of the measurement of belief, that has not yet been explored, but that may be promising for the future.

The basic idea is that it is a mistake to think of beliefs as being stored in some fixed and inert form in the memory of a person. When a question is asked about personal beliefs, one constructs a belief coded in a belief statement as a response to the question. As the kind of question varies, the construction varies, and the results vary. What I am saying about the construction of beliefs is similar to a view commonly held about memory, namely, that very little of memory represents an inert encoding. We are primarily constructing detailed memories by procedures that we do not at present understand, but that operate in a more subtle way on encoded data than simply by a direct retrieval of information. As many of you will recognize, such a conception of memory is classical and is especially associated with the early important work on memory by Sir Frederic Bartlett, especially in his book, *Remembering* (1932). A good recent overview of these matters, including an appraisal of the current status of Bartlett's ideas, is found in Cofer (1973).

Let me be clear about the basic point I want to make. After all, constructions are familiar in geometry and lead to exact results. A similar claim might be made about the constructive processes in memory in which we examine past experience in reaching comparative evaluations of belief. My point is, however, that the constructive processes in the case of belief are not of this kind, but are easily disturbed by slight variations in the situation in which the constructive processes are operating. This kind of view backs up the layman's view that it is ridiculous to seek exact measurements of belief; it can also be used to defend the expert opinion that it is unseemly to ask for the next decimal in a measurement of subjective probability. I do not at the present time have any good ideas of how to think about these constructive processes. My conjecture is that this is a move in the right direction, and that in making this move we should try to operate at an abstract level that will lead to specific results in explaining the felt uneasiness of any attempts to seek exact measurements of belief.

My one definite idea about such constructive processes is that mathematical models of learning provide a preliminary, simple schema. Modern rationalists of human thought sometimes seem to think that beliefs are changed simply by the use of Bayes's theorem, or at least in first approximation this is what happens empirically. And, ideally, this is what always should happen in the case of a rational man. There are many ingredients for considering this Bayesian idea a fantasy of reason, however, and I have on a previous occasion tried to state several of them (Suppes, 1966). Let me summarize the matter by saying that in many cases of change of belief it appears obvious that we cannot identify directly or indirectly the evidence on the basis of which the belief is changed, much less the relevant likelihoods or probabilities.

Simple learning models that work in first approximation, both for animals and humans, give some idea of how such constructive processes operate. It seems appropriate to say that the kind of changes that take place in learning can be regarded as examples of changes in belief. Thus if we study as a process of stimulus sampling and conditioning the acquisition of simple mathematical concepts by children, it is correct to say that during the course of learning, their beliefs about the concepts being taught change, as reflected in their responses. In making this remark, I am not suggesting for a moment that changes in belief are always reflected in responses, but rather that this is one way of getting evidence on changes of belief.

It is of course sometimes said that learning theories that postulate learning primarily on the basis of stimulus sampling and conditioning are too passive in nature, and that they do not consider adequately the conscious use of cognitive strategies by learners. I think that on occasion conscious strategies are used but, ordinarily, these strategies are not articulated, and when a learning theory based on stimulus sampling and conditioning is formulated in proper mathematical terms (see, for example, Estes, 1959; Suppes, 1969; Estes and Suppes, 1974), there is no commitment to whether the internal processes are constructive or passive in nature. The level of abstraction in handling the concept of stimulus is such that constructive processes could easily be assumed for handling the conditioning of stimulus patterns or, if you will, in more cognitive terms, the formation and storage of hypotheses.

To illustrate these ideas with a concrete but simple example I draw upon some earlier work reported in Suppes and Ginsberg (1963). The two-element model I consider may be conceptualized as follows. There are two stimulus features or patterns associated with each experimental situation. With equal probability exactly one of the two features is sampled on every trial. Let us call the features or elements σ and τ. When either element is unconditioned there is associated with it a guessing probability g_σ or g_τ as the case may be, that the correct response will be made when that unconditioned stimulus is sampled. An assumption of particular importance to the present model is that the probability of the sampled stimulus element becoming conditioned is not necessarily the same when both elements are unconditioned as it is when the non-sampled element is already conditioned. We call the first probability a and the second b.

Under these assumptions, together with appropriate general independence of path assumptions as given, for example, in Suppes (1969), the basic learning process may be represented by the following four-state Markov process, where the four states $(\sigma, \tau), \sigma, \tau$ and 0 represent the possible states of conditioning of the two stimulus elements.

	(σ,τ)	σ	τ	0
(σ,τ)	1	0	0	0
σ	$b/2$	$1-b/2$	0	0
τ	$b/2$	0	$1-b/2$	0
0	0	$a/2$	$a/2$	$1-a$

The model just described can be applied with reasonable success to data on children's learning simple mathematical concepts. A typical example would be the experiment on geometrical forms of Stoll (1962) reported in Suppes and Ginsberg (1963). In this experiment the subjects were kindergarten children who were divided into two equal groups. For both groups the experiment required successive discrimination, with three possible responses permitted. One group discriminated between triangles, quadrilaterals and pentagons, and the other group discriminated between acute, right and obtuse angles. For all subjects a typical case of each was shown immediately above the appropriate response key.

I shall not go into the detailed analysis of data; those interested are referred to the references just given. From the standpoint of concern here, it is easy to see why passive stimulus sampling seems absurd. The stimulus displays varied from trial to trial. For example, the same acute angle was not displayed on each trial. Obviously the subjects had to go through the constructive process of approximately matching the salient features of the display (conceptualized by the model to be two in number) to obtain a decision on their presence or absence. It is certainly true that the kind of theory I have described does not provide adequate details of this processing. It does provide a coarse analysis that fits data remarkably well. This is a simple example, but hopefully illustrates my point. The subjects were too young to verbalize in any precise way what they had learned, but they were able to learn the constructive processes of identification, and their beliefs and knowledge were changed in the process. There is a good deal more I would like to say about how these internal constructive processes operate. Conceptually I currently think of them in terms of computer programs written in terms of a simple set of instructions involving perceptual as well as internal processes. The features σ and τ in the simple model described above are each represented internally by two elementary programs. In a recent publication I have tried to spell out this approach to learning for the case of children's acquisition of the standard algorithm of numerical addition, but it is not possible to enter into detail here (Suppes, 1973a).

5. Final remark

When one examines the status of learning theory in relation to complex concepts, or the analysis from any other standpoint, including contemporary cognitive psychology, of the acquisition and holding of beliefs, it seems appropriate to be skeptical of our ever achieving a complete theory of such matters. The information we can obtain about an individual's beliefs will, in my judgment, always be schematic and partial in character. Even if the time comes when we shall be able to have what we feel is an adequate fundamental schema of the processes involved, it is doubtful that we shall be able to implement a complete quantitative study of an individual's beliefs.

To accept the necessary incompleteness of what we can analyze is, to my mind, no different from accepting the impossibility of complete meteorological predictions. It is hopeless and, probably in one sense, uninteresting to attempt to measure and predict exactly the motion of the leaves on a tree as a breeze goes by. Our beliefs, it seems to me, are rather like the leaves on a tree. They tremble and move under even a minor current of information. Surely we shall never predict in detail all of their subtle and evanescent changes.

15

THE LOGIC OF CLINICAL JUDGMENT: BAYESIAN AND OTHER APPROACHES

Not many years ago it would have seemed impractical, if not impossible, to have physicians and philosophers engaged in dialogue about the logic and nature of clinical judgment. The philosophers would have been unwilling or unprepared to think about matters that on the surface seemed far removed from classical philosophical problems. Physicians on their part would have been wary of entering into the labyrinth of methodological issues dealing with the relation between judgment and evidence. Now it seems wholly natural to have such an interaction and to have a conference that focuses on clinical judgment, with physicians and philosophers doing their best to interact and to understand each other's problems and methods.

I am sure that a difficulty for all of us is not to get carried away with expounding the technical subjects on which we are now working and to strive to communicate at the appropriate level of generality and simplicity. I know from experience that medical talk about any specialized area of disease can almost immediately get beyond my competence and knowledge if the full clinical details are presented. Over the past several years I

*Reprinted from *Clinical judgment: A critical appraisal* (ed. by H. T. Engelhardt, Jr., S. F. Spicker and B. Towers), 1979, pp. 145–159. Dordrecht, Holland: D. Reidel Publishing Co.

have had the pleasure of talking about matters that are generally relevant to this conference with my colleagues in the Stanford Medical School. A number of these conversations have been with members of the Division of Clinical Pharmacology. I have been pleasantly surprised at my ability to get a sense of the problems they consider important to attack, even though the detailed terminology and data of clinical pharmacology lie outside areas of knowledge about which I claim to have accurate ideas. I think that the same goes for my own areas of special knowledge. It would be easy enough for me to raise particular questions in the foundations of probability or decision theory that are of current concern to me and that have some general relevance to the theory of clinical judgment, but that would be too specialized and esoteric for detailed discussion in this context. No doubt I shall not be able to be totally austere in this forbearance and will occasionally at least allude to current technical interests of my own that have potential relevance to the topic of this conference.

There is of course another danger—a practice of which philosophers are often guilty—that what I have to say could be formulated in such a general way that it would not really be of interest to anyone, perhaps because the ideas in their most general form are already widely familiar.

With these considerations in mind I have divided my paper into four sections The first deals with probability and the general foundations of statistical inference, with attention focused on the Bayesian approach. The second section enlarges the framework of probability to that of decision theory by introducing the concept of the value or utility of consequences. Unlike many applications of modern decision theory to scientific research, the application to clinical judgment seems especially natural and appropriate. The third section deals with models. The main point here is that a general theory of decision making is no substitute for particular scientific understanding. The fourth and final section deals with what seem to be some of the perplexing problems of data analysis in medicine, at least from the perspective of an outsider who has had more problems and experience with data analysis in other areas than he cares to think about.

1. PROBABILITY

Among fundamental scientific concepts, that of quantitative probability is a late arrival on the scene, as my colleague Ian Hacking has shown in a splendid monograph on the emergence of the concept in the 17th century (1975). The theory of statistical inference is even more recent, and is really only a product of the 20th century. Given the long and developed history of medicine, reaching back for thousands of years, it is not surprising that the recent concepts of probability and statistics have as yet had little impact on the practice of medicine.

Moreover, some of the foundational views of probability do not in any natural way lend themselves to the clinical practice of medicine. One important and fundamental approach to probability has been to emphasize that probability always rests on the estimation of relative frequency of favorable cases to possible cases in some repeatable phenomena, of which games of chance provide paradigm examples. The long history of emphasis in medicine on the diagnosis of the individual case does not easily lend itself to this relative frequency view of probability.

Fortunately, there is an equally persuasive and important view of probability as expressing primarily degree of belief or, as it is sometimes put for pedantic purposes, degree of partial belief. In ordinary talk, most of us consider it sensible to ask what is the probability of rain tomorrow. We have even come to expect the evening TV news to provide a numerical estimate. When the forecaster says that the chance of rain tomorrow is 60 percent, he is not using in any direct way the relative frequency approach but is expressing his degree of belief even if he does not himself explicitly use such language.

Physicians, it seems to me, generally do not need any persuasion about the importance and value of the expression of a degree of belief as an approach to probability. This approach is often called Bayesian because of its early lucid formulation as a foundational viewpoint by the Reverend Thomas Bayes in the 18th century (1763).

The centerpiece of this approach is Bayes' theorem, which says that the posterior probability of a hypothesis, given evidence, is proportional to the likelihood of the evidence, given the hypothesis, times the prior probability of the hypothesis itself. If consequences are ignored, then the maximum of rationality that follows from Bayes' theorem is that we should act on the hypothesis that has the highest posterior probability.

Let us examine some of the difficulties in a direct application of Bayes' theorem to clinical practice. A general problem is the unwillingness of many physicians, on the basis of temperament and training, to put themselves in an intellectual framework that calls for probability judgments in diagnosing a patient's illness. It seems to me that there are two good intellectual reasons for this resistance on the part of physicians. The first is skepticism that a mechanical or semimechanical algorithm can be as effective in assessing a patient's state as the intuitive judgment of an experienced diagnostician. It has not been part of the long tradition of clinical practice to attempt numerical assessments of any kind, really, of a patient's state, and an experienced clinician can be skeptical that a reduction to a numerical statement is feasible. There are many kinds of decisions or judgments we make that are not easily or naturally reduced to verbal rules, let alone quantitative rules. Perhaps one of the simplest examples that has received a good deal of study in experimental psychology is the way in which we recognize faces or familiar smells. Either in the visual case of face recognition or in the olfactory recognition of familiar smells the verbal descriptions we can give of the evidence on which our decisions of recognition are based are extremely poor and vague in character. A reduction to explicit rule of recognition procedures in either of these relatively simple but familiar domains would probably be unworkable for even the most articulate. Similarly, so the argument goes, the intuitive judgment of clinicians is based upon a depth and range of experience that cannot be reduced to explicit rules.

The second major argument against Bayes' theorem is that even in arenas where explicit data, for example, of a laboratory sort, are being considered and a framework of explicit concepts is being used, there is often no natural and nonarbitrary way to incorporate new objective evidence within a feasible application of Bayes' theorem. In this case the evidence is explicit and the data are objective, but we do not have explicit rules for calculating likelihoods. We especially do not have such rules when we suspect there is strong probabilistic dependence among various parts of the evidence being considered.

I respect both of these arguments in an essential way. As far as I can see, there is no reason to believe that the time will ever come when we can have any simple direct mechanical application of Bayes' theorem or similar statistical tools to provide a satisfactory but automatic diagnosis of an individual patient's illness. This does not mean that I am against pushing the use of Bayes' theorem and other similar methods as far as we can and indeed insisting on as many studies as possible of their feasible application. A number of studies, in fact, of the use of Bayes' theorem in medical diagnosis have already been made, and several with quite positive results. For example, Warner et al. (1964) incorporated Bayes' theorem in a computer program that was used in the diagnosis of congenital heart disease. The program that applied Bayes' theorem classified new patients with an accuracy close to that of experienced cardiologists. I shall not attempt to survey here the number of other excellent studies in this direction. A good brief survey is to be found in Shortliffe (1976), who is sympathetic to what has been demonstrated thus far but who is at the same time dissatisfied with a Bayesian statistical approach as being anything like the final word.

The remaining three sections of this paper deal with broad concepts that I think are necessary to augment a Bayesian approach to clinical judgment. Before turning to these matters, I do want to emphasize that I have made no attempt here to enter into the deeper technical developments of the Bayesian theory of statistical inference or alternatives that have been extensively studied by mathematical statisticians over the past several decades.

2. EVALUATION OF CONSEQUENCES

The standard Bayesian application often emphasizes only the assessment of beliefs and how these beliefs change with the accumulation of new evidence. The general theoretical setting, however, of decision theory has emphasized the fundamental place not only of belief but also of evaluation of consequences. It is my impression that the current literature on clinical judgment in medicine has placed much more emphasis on methods for assessing beliefs consistently and rationally than it has on assessing rational methods of evaluating the consequences of the decisions taken. There is undoubtedly a sound intuition back of this emphasis. If all the evidence is in, it often seems clear enough what action should be taken and what the anticipated consequences of the action will be. For example, if a patient is diagnosed with extremely high probability to have a particular infectious disease for which there is a standard, highly specific treatment, and, moreover, the probability is low that any known side effects of the treatment will have deleterious effects on the patient, then the main task of decision making is over. The consequences of deciding to prescribe the standard treatment seem obvious and do not require extended analysis.

The difficulty, of course, is that this kind of clear situation seldom obtains. Moreover, in the context of modern medicine, a new factor has arisen which has already led to some emphasis being given to problems of evaluation of consequences. This is the problem of holding down the cost of laboratory or physical tests (Gorry and Barnett, 1967-1968).

In practice, of course, physicians do automatically attach some evaluation of consequences, including the cost of laboratory tests, for if they did not, the rational decision would always be to require as many laboratory tests as possible in order to maximize the evidence available in making a diagnosis. Reasonable rules of thumb no doubt were appropriate and proper in the past. With the much more elaborate possibilities available today, and with the costs rapidly rising of the more sophisticated laboratory tests, it also seems appropriate that more elaborate tools of decision making will come to have a natural place in clinical judgment.

It is safe to predict that, with the national concern to control the costs of health services, attention to the problem of costs just mentioned will be a major focus of both theoretical and practical work on clinical judgment. I would like, however, to focus on some different issues that arise from the evaluation of consequences and that have possible implications for changing the traditional relation between physician and patient. What I have to say about these matters is certainly tentative and sketchy in character, but the issues raised are important and, moreover, the tools for dealing with them in a rather specific way are available.

The issue I have in mind is that of making explicit the possible consequences of a decision taken about medical treatment on the basis of clinical judgment. Traditionally, no very explicit model of evaluation is used by the physician either for his own decision making or for his consultation with the patient about what decision should be made. It is a proper part of the traditional relation between physician and patient that with certain unusual exceptions, the final decision about treatment is the patient's and not the physician's. On the other hand, it is a part not only of traditional but also of modern medicine for the vast majority of patients to accept the treatment that is preferred by the physician. I have not seen any real data on this question but it would be my conjecture that in most cases the physician makes relatively obvious to the patient what he thinks is the preferred treatment. In cases of certain risky operations or experimental drug treatments, etc., almost certainly there is a much stronger tendency to lay out the options for the patient and to make explicit to him the risks he is taking in the decisions he makes.

It is especially the decisions that have possibly grave negative consequences to the patient that suggest a more explicit analysis of the decision process. To provide a concrete example for discussion, let us consider the following highly simplified case. I hope that you will bear with the obviously oversimplified character of my description. Let us suppose that the patient is one with a serious heart condition. He is presented by the heart surgeon to whom he has been referred with any one of three options: bypass surgery, continual treatment with drugs without surgery, no treatment of any sort. To carry through an explicit decision model of a quantitative sort, the patient needs now to be presented data on possible consequences of each of the three actions, together with the medical evidence on his heart condition. According to the standard expected utility scheme, the patient should then select the medical treatment that will maximize his own expected value or expected consequences or, in other terminology, expected utility.

Satisfaction of the conditions required to make an expected-value model work well do not seem easy to come by even in the most clearcut clinical situations, and consequently I would like to examine in some more detail the feasibility of using such a model at all.

The problems that arise naturally separate into two parts. One part concerns the ability of physicians to make quantitative assessments of possible consequences of treatment as well as of the current true state of the patient as inferred from the evidence available. We can question whether the state of the art and of the science can bear the load of such quantitative assessments at the present time. We may certainly want to hold the view that assessment of individual cases in such a quantitative fashion is not practical. I will return to this point in a moment. The second part concerns the ability of the patients to absorb the data concerning the options presented to them by the physician. The number of patients who feel at home with probability calculations or the concept of expected value is very small indeed. Anything like a routine quantitative application of the model seems totally impractical for all but an extremely small segment of patients, at least at the present time. It may properly be claimed that the detailed quantitative assessment of evidence or of consequences is as complicated a technical topic as the laboratory tests called for by the clinician, and the ordinary patient is simply not competent to deal with a quantitative decision model even when its application is of great personal consequence to himself.

There is a second approach that seems a good deal more promising in the present context of medical practice and the expected knowledge of patients. This is to move the development and analysis of a quantitative decision-making model from the level of the individual case to statistical analysis of a large number of cases. It is certainly true, for example, that the consequences will vary enormously from one patient to another, not only because of his physical condition but also because of his age, his wealth, his family responsibilities, etc. On the other hand, there are clearly four consequences that dominate the analysis of the full nexus of consequences, namely, (i) the probability of recovery, (ii) the probability of death, (iii) the probability of serious side effects in terms of medical consequences, and (iv) the expected cost in terms of types of treatment. In summary, the direct medical consequences and the direct financial costs of a given method of treatment are the most important consequences, and these can be evaluated by summing across patients and ignoring more detailed individual features. This does not mean, for example, that in assessing the consequences of treatment we ignore the age of the patient, because this is part of the evidence and should go into the assessment of the consequences for the given state of the patient.

The unconditional or mean assessment of the consequences of particular treatments is a relatively straightforward piece of data analysis. The situation is quite different, however, if we want to make the appropriate conditional assessment—conditional upon the variation in relevant parameters of the patient's state at the time of treatment. The complexity of estimating the joint probability distribution of various consequences—or symptoms—is admitted by almost everyone who has considered the problem.

What seems desirable at the present time is development of an actuarial approach to both the state of health and the consequences of treatment. Such an actuarial analysis could serve only as a guideline in the treatment of individual cases, but useful information about medical decisions by individual physicians or groups of physicians could be obtained.

Let me give an example. A number of studies (Peterson, Andrews, Spain and Greenberg, 1956; Scheckler and Bennett, 1970; Roberts and Visconti, 1972; Kunin, Tupasi, and Craig, 1973; Simmons and Stolley, 1974; and Carden, 1974) have shown that there is a definite tendency for nonspecialists to prescribe more antibiotics than are required by patients' conditions. Studies of this kind provide an excellent way of cautioning physicians to consider carefully the clinical basis of any prescription of antibiotics but do not attempt at all to provide an algorithmic or mechanical approach to clinical diagnosis.

3. MODELS

Although the merits of Bayesian and related methods of inference can be defended as practical tools that can be brought to bear on real problems of clinical judgment, it is important to emphasize that such methods are no panacea and do not provide in themselves a scientific foundation for medicine that is in any sense self-sufficient. It is quite true that there are areas of medicine that are clinically important and that do not at present have a thoroughly developed theory. My better informed friends tell me that this is true of more areas of clinical medicine than I would be naturally inclined to believe, but I certainly won't venture to give details on this point and simply take it as an assumption that it is easy to draw distinctions between various areas of clinical medicine. The distinction is concerned with those that have a practically applicable theory and those that do not. Those that do not, it seems to me, can especially benefit from the application of Bayesian methods, but this benefit should not obscure the need to continue the development of a more adequate scientific foundation. Moreover, the development of that better scientific foundation surely does not depend on any direct application of statistical methods, but rather on the creative development of new scientific concepts and theories. Explicit scientific models of the relevant biological phenomena must remain a goal, I would suppose, in every area of clinical medicine.

As such models develop, we should be able to fold them into a general framework of statistical inference and there should be no natural conflict between the use of newly developed models and the stable data-based inferences of the past. There will, of course, be the practical problem of assaying the relative weight to be given to past experience, on the one hand, and, on the other, the relative weight to be given to the new scientific models of the phenomena at hand.

There is nothing about this situation that is distinctive and special to the problems of clinical medicine. A similar tension between past experience and the urge to develop deeper scientific models is to be found in every area of applied science. Salient examples that confront us every day and that are nearly as important as the problems of medicine are to be found in economics and meteorology. Moreover, we are all conscious of the difficulties of developing adequate scientific models either of the economy or of the weather, but the thrust to do so is deep and sustained, just as it is in modern medicine.

The introduction of more general and more powerful scientific models in various clinical areas seems to me to generate an interesting and important problem of differentiating the future possibilities. On the one hand, the scientific thrust is to make the clinical diagnosis of a patient ever more algorithmic. Although, as I have argued above, we shall never pass beyond the need for clinical judgment, it is still important to recognize that what we can see in the reasonably near future for different parts of medicine presents quite a different picture. In the diagnosis of infectious diseases, for example, we might expect to get nearly algorithmic laboratory and computer-based procedures—at least I will venture that conjecture. On the other hand, in the diagnosis and treatment of psychiatric disorders we may anticipate in no reasonable future a scientific model of sufficient depth and generality to provide anything like algorithmic diagnostic procedures. I wish that I were competent to give a survey of the various areas of medicine and to conjecture what we might expect along algorithmic lines. I would be enormously interested in hearing more informed opinion than my own about this matter.

It is also not clear what we may expect from models that derive from work in artificial intelligence. The MYCIN program of Shortliffe (1976), for example, has many attractive features and is potentially a diagnostic aid of great power, but it is still rather far from being ready for practical daily use, even in the sophisticated setting of a teaching hospital.

4. DATA ANALYSIS

Let me begin my remarks about data analysis with a tale of my own. A couple of years ago we embarked on collecting a large corpus of spoken speech of a young child. The mother of the child was hired as a half-time research assistant to spend twenty hours each week creating a properly edited computer file on which she transcribed an hour of the child's spoken speech for that week. In something over a year of effort on the part of the mother we obtained a corpus of more than a hundred thousand words of the young child from the age of approximately two years to three and a half years. We then engaged in elaborate computations regarding structural features of the speech, especially an elaborate test of a number of different generative grammars and model-theoretic semantics and of developmental models of grammatical usage. We estimated that at the conclusion of our elaborate computational analysis of a number of different grammatical models we had probably done more explicit computing than had been done by all the linguists of the 19th century working on all the languages examined. Even so, the piece of data analysis we did on a child's speech seems trivial compared to the overwhelming problems of data analysis in clinical medicine. In just one large clinic, consisting of, say, a hundred doctors and their support staff, the data flow is like a torrent and the problem of providing sensible analysis appears almost overwhelming.

However, it seems to me that there is much that is constructive that can be done and that can provide important supporting analyses for those responsible for final clinical judgments.

The first fallacy to avoid in attempting this is the philosopher's fallacy of certainty. There is no hope of getting matters exactly right and organizing a body of data that will lead to certain and completely reliable conclusions about any given patient. It is important to recognize from the start that the analysis must be schematic, approximate, and in many cases crude.

Second, the penchant of many social scientists and applied statisticians for experimental designs must be reçognized in this context as a romantic longing for a paradise that can never be gained. Just because it is not practical to impose experimental methods of design or parameter variation on the flow of patients through a standard medical clinic, it does not follow that any quantitative approach to causal analysis must be abandoned. Interestingly enough, some of the very best modern methodology has been developed and is being used by econometricians dealing with data that are totally inaccessible to experimental manipulation. Moreover, the statistical analysis of data was first used in a massive way in the least experimental of the physical sciences, namely, astronomy. We should no more despair of the severe limitation on experimentation in clinical medicine than astronomers of the 18th century despaired at the absence of the possibility of astronomical experiments. I want to make this point explicit because it is one that I have spent a good deal of time on in casual argument with philosophers and statisticians whose opinions I respect but do not agree with. One of the most sophisticated and significant applications of probabilistic and statistical analysis to the identification of causes was Laplace's method of what he termed "constant" causes. He used these methods to attack the subtle problems involved in the effect of the motion of the moon on the motion of the earth, to analyze the irregularities in the motion of Jupiter and Saturn, and to identify and consequently to explain the mean movements of the first three satellites of Jupiter. These are classical results from the late 18th century, but one has to look far and wide in the entire history of science to find *experimental* results of comparable conceptual and quantitative sophistication.

The third point amplifies some earlier remarks about joint probabilities. Even crude approximations to the joint distribution of causes or symptoms would, I would conjecture, be a definite methodological step forward in the analysis of clinical data. In this case I am returning to my earlier remarks about looking at large numbers of cases and applying the results as guidelines for considering individual patients. It is a first lesson in elementary probability theory that from the (marginal) distributions of single properties it is not possible to infer the joint distribution of the properties. The estimation of these joint distributions is a complex and subtle affair, but it is my belief that in many cases even relatively crude results would lead to clinical insights of considerable interest.

The fourth point concerns the great importance of considering alternative hypotheses or causes to provide a perspective on the identification of the most likely cause. Consideration of alternative hypotheses is natural within a Bayesian or a classical objective framework of statistical inference and is a matter that is old hat to statisticians, but it is not only in medicine but in other parts of science as well that the explicit formulation and analysis of the data from the standpoint of alternative hypotheses are far too often not undertaken. I do not mean to suggest that good clinicians do not range over a natural set of possibilities in diagnosing the illness of a patient, but rather that, in the kind of quantitative data analysis based on many cases that I am advocating as a general intellectual support, analysis is often not adequately presented of the support the data give to alternative hypotheses or causes.

To show that the philosophical thrust of this last remark is not new, as indeed is true of most of the other things I have had to say, let me close with a quotation from Epicurus's letter to Pythocles, written about 300 B.C. soon after the very beginning of philosophy as we know it. Epicurus's remarks are aimed at our knowledge of the heavens or, more generally, of the universe around us but they apply as well to the focus of our present discussion.

> For this is not so with the things above us: they admit of more than one cause of coming into being and more than one account of their nature which harmonizes with our sensations. For we must not conduct scientific investigation by means of empty assumptions and arbitrary principles, but follow the lead of phenomena: for our life has not now any place for irrational belief and groundless imaginings, but we must live free from trouble. Now all goes on without disturbance as far as regards each of those things which may be explained in several ways so as to harmonize with what we perceive, when one admits, as we are bound to do, probable theories about them. But when one accepts one theory and rejects another, which harmonizes just as well with the phenomenon, it is obvious that he altogether leaves the path of scientific inquiry and has recourse to myth (Oakes, 1940, p. 11).

16

ARGUMENTS FOR RANDOMIZING

I have organized my remarks about randomizing under four headings: computation, communication, causal inference, and complexity. It is hard to think of a more controversial subject than that of randomization. My remarks are simpler and more extreme than they ought to be. I have put them in a rather bald and definite way in order to draw the lines more sharply and to make my message as clear as possible. I do not doubt that under extended debate it would be necessary to qualify some of the things I have to say, but I would insist on the point that I would be offering qualifications, not retractions.

1. COMPUTATION

It is often said by pure Bayesians that once the likelihood function is available knowledge of any randomization scheme used is superfluous information. It seems to me that this argument misses an important point which I want to illustrate by a simple artificial example.

Suppose I am presented with an urn in which I am told that there are fifty balls and the mixture of white and black balls satisfies one of two hypotheses. The first hypothesis is that there are fifteen black balls and

*Reprinted from *PSA 1982*, Vol. 2 (ed. by P. D. Asquith and T. Nickles), 1983, 464–475. East Lansing, MI: Philosophy of Science Association.

thirty-five white balls. The second hypothesis is a symmetric image of this one, namely, fifteen white balls and thirty-five black ones. I am now told that I can draw with replacement a dozen balls, and on the basis of the outcome of the twelve draws state which hypothesis I would bet a hundred dollars on at even odds. (I have to be willing to make the bet in order to participate in the experiment).

I do not know about some Bayesians (I count myself a semi-Bayesian) but I certainly know what I would do in this situation. I would insist on a thorough mixing of the physical position of the balls in the urn. I would want to supervise this physical mixing myself and I would want it done in such a way that I believed I had approximated a uniform distribution (what I mean by uniform distribution here is clear enough from the compact character of the container) for the location of any ball in the urn. I find it hard to imagine a sophisticated bettor who would not insist on such physical randomization before entering into the experiment. Without such randomization I would not be able to write down the standard likelihood function under each hypothesis for the twelve draws. Why? Because the likelihood function I believe would depend on the physical distribution of the balls in the urn. If I did not physically randomize the positions of the balls, I would use a prior with high variance over the possible physical distributions. I might feel that I could compute a likelihood function based on this prior but I would be uncomfortable doing so. Given that I could reduce the variance on my prior enormously by insisting on physical randomization, I would strongly insist on doing so. What I have proposed to do here is go from a prior to a posterior in terms of the *physical* distribution of the balls in the urn. This is a different posterior than my posterior based upon the likelihood function used to compute the posterior distribution after drawing the twelve balls. Note also that the posterior distribution concerning the physical distribution of the balls in the urn is one I achieved without going through the step of moving from a prior to a posterior via a likelihood function. Moreover, since I am sampling with replacement, this act of randomization takes place after each draw and is what justifies the independence assumption in the sampling that makes the likelihood hypothesis so direct and easy to compute.

Because the main attack against randomization in experimentation has come from Bayesians and because I myself accept very much of the Bayesian viewpoint, it seem particularly desirable to justify randomization from a Bayesian viewpoint. I take it that it will be more or less accepted that statisticians of the Neyman-Pearson or Fisher type require little if any justification. For Bayesians, the physical randomization described has two principal virtues. First, it provides an alternative to sampling, which is not available, for passing from a prior with a very high variance on possible distributions of physical location to what is essentially a single distribution. The importance for Bayesians of physical randomization to change distributions or, if you will, to reach a posterior distribution by other means than sampling has not been adequately recognized. I have pointed out elsewhere on several occasions that it is a paradox of Bayesian thought that random sampling with the computation of a posterior distribution via a standard likelihood function is the single most natural way of gaining information for Bayesians. It is difficult for Bayesians to incorporate changes in their own opinion or information in the form of opinion of others into their prior distributions. (For an extended example see Suppes, Macken, and Zanotti, 1978.) Physical randomization is a fundamental and important method for Bayesians for fixing the distribution used in computing the likelihood function.

The second principal virtue for Bayesians of the process of physical randomization is that it is a method of introducing a distribution that makes possible *simple* computations. It is also possible to make a stronger claim. If physical randomization is not used it is not clear how to incorporate into the likelihood function qualitative information that might be part of a Bayesian's beliefs about how the balls were put in the urn originally. It is a common view of Bayesians that priors that incorporate some qualitative aspects of beliefs only rather crudely are not a problem, because extensive sampling will make such discrepancies unimportant, but when such beliefs center around the likelihood function the problem is potentially serious and cannot be ignored. Bayesians find it easy to disagree about prior distributions but troublesome to disagree strongly about likelihood functions because when likelihood functions are different it is not possible, in general, to get convergence of opinion with sampling. Bayesians can live with such a state of affairs but it is important to emphasize how fundamental in practice an agreed-upon likelihood function is. This agreed-upon likelihood function is often due, as in the present case, to a process of physical randomization. Such physical randomization can be an important component of many kinds of experiments.

The principles involved in this simple and artificial example apply, with the expected complications, to a wide variety of real examples where randomization is current standard practice. On the other hand, when a well-defined theoretical model is postulated, randomization is often not required. Thus I want to make clear that I do not think that randomization is an intensive feature of every possible kind of experimentation. Let us first consider an example in which it is not.

An example in which randomization is not required is provided by an ergodic Markov chain whose transition probabilities depend upon a single real parameter, say θ. By saying that it is an ergodic Markov chain we mean the following things, intuitively speaking: (1) the trials are discrete rather than continuous in nature; (2) the probability of being in a given state on trial n depends only on the state on trial $n-1$; (3) the number of states is finite; (4) the transition probabilities are independent of time or trial number; and (5) the probability of being in state k on trial n is in the limit as the number of trials increases independent of the initial state, i.e., the state on trial 1.

It is often not difficult to obtain the maximum-likelihood estimate of θ. Such an estimate then provides us all that we need to know in order to test the validity of the postulated Markov chain as the theory of the phenomenon being observed. The essential point is that randomization need not enter in any essential way in conducting the experiment that tests the theory. We can observe, as is often the case in many different kinds of experiments, a single sample path. Even the ergodic character of the chain, this single sample path is theoretically adequate as an empirical test of the theory. The theory itself then gives us the assumptions about the likelihood function needed to make actual computations and find the estimate of θ. In particular, let $a_1, a_2, \cdots, a_n$ be the initial segment of n trials of a sample path of the process. Then the maximum-likelihood estimate of the learning parameter θ is the number θ (if it exists) such that for all θ

$$f(a_1, a_2, \cdots, a_n; \widehat{\theta}) \geq f(a_1, a_2, \cdots, a_n; \theta). \tag{1}$$

Here $f(a_1, a_2, \cdots, a_n; \widehat{\theta})$ is the probability of the sequence of responses $a_1, a_2, \cdots, a_n$ when the learning parameter is $\widehat{\theta}$.

By virtue of the fundamental Markov property of the process, we have

$$f(a_n|a_{n-1};\theta)f(a_{n-1}|a_{n-2};\theta)\cdots f(a_2|a_l;\theta)f(a_1;\theta) = f(a_n, a_{n-1}, \cdots, a_1;\theta). \tag{2}$$

Now as an approximation we ignore the probability of the state on the first trial and look only at the transitions. This is a reasonable approximation when the number of trials is large. So, summing over trials we want to maximize

$$\prod_{m=2}^{n} f(a_m|a_{m-1};\theta). \tag{3}$$

Let N be the number of states in the process; $p_{ij}(\theta)$ the probability of going from state i to state j with parameter value θ; n_{ij} the observed number of transitions from state i to state j aggregated over trials—and so the n_{ij} are our experimental data. Substituting this notation in (3) we then replace (1) by (4).

$$(4) \qquad \prod_{i,j=1}^{N} p_{ij}^{n_{ij}}(\hat{\theta}) \geq \prod_{i,j=1}^{N} p_{ij}^{n_{ij}}(\theta).$$

A variety of psychophysical experiments exemplify the kind of application just discussed. It is characteristic of such experiments that usually more than one subject is used but often no more than five or six, and each of these subjects stays in the experiment for a very large number of trials, for example, 10,000 would not be unusual. Each subject is often treated as an independent realization of the theory and a separate parameter is estimated. In no essential way does randomization enter in the selection of subjects. (In other ways, randomization can enter, for example, if the presentation of stimuli is probabilistic in character, but I am not considering that aspect in the present discussion.) From the standpoint of concern at the moment, the point is that randomization is not needed in order to provide a well-defined basis for computation, because that is provided already by the underlying theory being tested. For those who are unhappy with calling such a single Markov process a theory, I should mention that in the happiest of situations the transition probabilities of the process would be derived from general qualitative postulates of the theory and not simply be baldly assumed. This is especially true when the form of the transition probabilities as a mathematical function of the parameter is rather complex. We ordinarily are not satisfied if there is not some kind of intuitive derivation of such complex expressions from relatively simple and plausible assumptions about behavior of phenomena in the domain under investigation. Notice, of course, that if we want to make a strict inference about a given population then we would take the further step of random sampling. For example, if we were studying some psycho-acoustical phenomena and wanted to conclude with an inference about the population of adults with normal hearing we might very well want to concern ourselves with taking an appropriate random sample of such adults. In ordinary scientific practice in psycho-acoustical work such inferences are not made. Rather, detailed studies are made of individual subjects and the methodology is similar to that to be found in physics where one does not sample, for example, in any agreed-upon fashion from a population of particles, moving bodies, etc., according to an explicit sampling scheme.

The second example is conceptually the opposite of the one just given. Instead of having one sample path extending over a large number of trials, a large number of individuals are sampled and parameters of the distribution are estimated. Here the classical methodology both of Bayesians and non-Bayesians is to use random sampling, and the reason is straightforward; the likelihood distribution is based upon the assumption of random sampling. I emphasize the following point in as explicit terms as possible. The assumption of random sampling replaces and plays the role in the likelihood function that the theoretical assumption of a Markov chain played in the previous example. Let me give a simple example. Suppose my prior distribution is a Beta distribution on the presence or absence of a given property in a given environment. As a good Bayesian I want to draw a sample and obtain my posterior distribution. The absolutely standard way to do this for Bayesians is to draw with replacement a random sequence of n objects and obtain thereby a posterior distribution that is also a Beta distribution. The details are as follows and are very similar to those given above but, as I continue to emphasize, the likelihood function depends upon the random sampling.

Let $B(a, b)$ be the prior Beta distribution. We now draw a random sample of size n with constant probability θ of success, i.e., presence of the property. Let r be the number of successes. Then the posterior distribution of θ is $B(a + r, b + n - r)$. Here is a brief indication of the proof. The prior Beta distribution is:

$$f(\theta) = \frac{(a+b+1)!}{a!\,b!}\,\theta^a(1-\theta)^b,$$

where a, b are nonnegative integers. The likelihood, when $a_1, \ldots, a_n$ is the sample drawn is:

$$f(a_1, \ldots, a_n, \theta) = \theta^r(1-\theta)^{n-r}$$

and so the posterior is:

$$f(\theta|a_1, \ldots, a_n) \approx \theta^{a+r}(1-\theta)^{b+n-r},$$

where the coefficient has been omitted.

To replace this simple and straightforward scheme with nonrandom systematic sampling is not impossible by any means, but it is not obvious what is the most natural way to proceed, including construction of the likelihood function. The weakness of the alternatives is one of the best current arguments for randomization, especially weakness at the level of technical implementation.

2. COMMUNICATION

It can be conceded to "pure" Bayesians that in many situations the optimal experimental design relative to the prior held by the Bayesian is a deterministic rather than random design. The recent literature of mathematical statistics is full of examples—current working practice is another matter. In any case it is not irrational of statisticians to recommend to their scientific clients, or irrational on the part of the scientists to continue to use, random designs. One argument is that in designing an experiment one wants to present a design and results that are meant to be persuasive, above all, to others. It is a favorite maxim of many philosophers of science that one of the most important things about science is the continued critical review of results, both theoretical and experimental, by the community of interested and competent workers in a given domain. It is for this community that the scientist will use random designs rather than deterministic ones. In the case of more applied work the community is broader but the reasons for randomizing are the same.

There is, it seems to me, lurking in the background of such discussions a skeptical view that is not often enough put on the table. If an experimenter uses a deterministic design that is optimal from the standpoint of his Bayesian distribution, it is not unlikely that some competitive or critical fellow worker will claim that the design was biased toward the desired experimental results. It is a favorite point in some experimental sciences to emphasize that well-designed experiments reporting null results are seldom published. There is, it is claimed, a natural inclination to bias the design toward obtaining favorable results. It can even be said, as it often is, that the bias can be unconscious on the part of the experimenter. Now the skeptic may reply that an experimenter who will do this will also fudge on the application of a proper random design. Yet it seems to me that it is exactly the difference between these two cases that is important. An experimenter who announces that he has used in his design a particular random scheme must take a very deliberate and conscious step amounting to a form of scientific perjury to deliberately violate the scheme he has chosen in order to bias the design toward the experimental outcome he favors. Scientists who may have let their unconscious biases go to work on a deterministic scheme will be stern with themselves and their coworkers in deliberately violating a random design that itself has been deliberately chosen. To put the matter in theological terms, permitting an unconscious bias to creep into a deterministic sampling scheme is a venial sin, but to deliberately violate a chosen random scheme is a mortal one.

There is, apart from questions of unconscious bias, still another argument from the standpoint of social communication for using random designs, namely, their universality and therefore ease of communication. Deterministic optimization of design with respect to a given prior distribution requires the elucidation of details that may, in fact, be tedious and uninteresting to absorb. Selection of a standard random design communicates at once to others the experimental design used and it is then possible to move quickly to the main substantive results in the study. It is also important in this connection to stress that if we use the criterion of minimizing least squares, for example, the difference between an optimal deterministic design and a straightforward random design will in general be small and, compared to other parameters of importance in experimentation, scarcely something to worry about. This ease of communication and the relative robustness of many standard random designs are important social arguments for randomization.

3. CAUSAL INFERENCES

A classic claim about randomization in experimentation has been that it guarantees the statistical validity of causal inferences. Since Fisher's heyday in the 1920s this has been the *raison d'etre* of the use of randomization in experimental design. It may be too strong to claim that randomization guarantees validity of a causal inference, but the safeguards randomization introduces are powerful ones and not easily replaced by Bayesian deterministic alternatives. My purpose is to amplify and defend this standard claim within an explicit causal framework and against some standard Bayesian criticisms. I use the concepts of prima facie, genuine and spurious causes introduced in Suppes (1970) and for simplicity restrict discussion to events rather than random variables, which are necessary for a probabilistic treatment of quantitative causal effects. I use capital letters for events and lower case t's as subscripts to denote the time of occurrence of events. (The problem of events that occur over an extended period of time is ignored.) In these terms, the event $B_{t'}$ is a *prima facie cause* of event A_t if and only if

(i) $t' < t$,
(ii) $P(B_{t'}) > 0$,
(iii) $P(A_t|B_{t'}) > P(A_t)$.

The occurrence of an event like $B_{t'}$ that raises the probability of A_t occurring is sometimes said to be simply a predictive or diagnostic event in the absence of a more substantial causal structure. It seems to me that the term *prima facie* is better than either *predictive* or *diagnostic*; spurious and genuine causes become the important special cases of prima facie ones.

An event $B_{t'}$ is a *spurious* cause in sense one of A_t if and only if B_t is a prima facie cause of A_t and there is a $t'' < t'$ and an event $C_{t''}$ such that

(i) $P(B_{t'}C_{t''}) > 0$,
(ii) $P(A_t|B_{t'}C_{t''}) = P(A_t|C_{t''})$.

A stronger definition is to require a partition of the sample space such that for *all* elements $C_{t''}$ of the partition

(i) $P(B_{t'}C_{t''}) > 0$,
(ii) $P(A_t|B_{t'}C_{t''}) = P(A_t|C_{t''})$.

For our purposes here the notion of spurious cause in sense one will be the easiest and most direct to apply. In using sense one I have relaxed the definition from the earlier formulation so as not to require an increase in the probability of A, that is, I have eliminated the condition that $P(A_t|B_tC_{t''}) \geq P(A_t|B_{t'})$. The reason for this is that we are often interested in cases in which the probability is actually decreased.

It is a familiar fact of statistical analysis that if I can look at the data after an experiment then I can always find an event $C_{t''}$ that negates the effect of a prima facie cause unless that prima facie cause is a sufficient cause, namely, produces A_t with probability one. In classical statistical methodology such looking at the data after the fact is considered inappropriate and even for Bayesians it is a procedure that must be handled with care. This is an important point and I want to be clear about my view of it. It is certainly not inappropriate to look at the data after an experiment or an observational study, and to use the data to conjecture new hypotheses. This is an important part of scientific methodology in moving from one hypothesis to another or from one theory to another, but the explicit and artificial construction of an event C by simply enumerating favorable or unfavorable cases is not appropriate and certainly is not sanctioned by Bayesians.

It is also obvious that randomization is no guard against the construction of such an event C, if the probabilities in question are not one. We may construct a sample by looking at the data, which gives us a conditional probability that nullifies the prima facie cause and makes it spurious. Randomization, on the other hand, does provide a strong safeguard in terms of selection of events prior to observing outcomes. It provides a means of averaging and thus guaranteeing that the conditional probability which we are examining is indeed the appropriate conditional mean. I want to expand on what I mean by the phrase "appropriate conditional mean". The first point is that we might be able to name an event prior to the experimentation such that if we conditionalize on it together with the prima facie cause then we would expect an effect, perhaps an effect of this additional conditionalization strong enough to nullify the prima facie cause. For example, in a medical experiment we might reasonably conjecture that individuals who had no cases of the disease under investigation in the three previous generations would show a null effect for the prima facie cause of prevention because the disease itself had such a low incidence in this sub-population. This is not surprising but is of course a result that deviates from our mean result.

The second point is that deviations from the conditional mean can be expected in small samples, but not when the random samples are sufficiently large. Much of the instructive and entertaining dialogue in Basu (1980) depends on the random samples being relatively small. Basu's "scientist" confesses he did not randomize the allocation but rather tried hard to strike "a perfect balance between the treatment and the control groups." He would, I am sure, have found his intuitive balancing much more difficult if 200 or 500 pairs rather than 15 had been the sample size to deal with.

This second point is also important for an issue in the theory of causality: can causes be characterized by purely probabilistic concepts? The definitions given above would suggest a simple affirmative answer, but this would be mistaken—as mistaken as saying that a stochastic process has a purely probabilistic characterization. The derivation of the laws or equations of most if not all scientifically interesting stochastic processes will depend on a variety of substantive empirical assumptions that may be couched in probabilistic language but are evidently not definable in purely probabilistic terms. A Markov assumption for a physical process embodies, for example, rich hypotheses about the constitutive structure of matter. In other words, the concepts used to define events or random variables are not reducible, but must be brought in from the outside, so to speak. Given, however, the concepts we are prepared to consider, random sampling with sufficiently large sample sizes can guarantee within the given framework of concepts to test correctly for any hypothesized genuine causes as well as prima facie ones.

It is sometimes maintained by certain methodologists holding an extreme position about the necessity of experimentation, that without randomization no causal effects can be discovered. Certainly I want to resist this claim, as the Markov chain example of the previous section makes evident. More generally, it seems patently false because we hold all kinds of causal beliefs about the world around us that have not been established by random experiments. Furthermore, in disciplines as distinct as astronomy, economics, and meteorology, valid causal claim are supported by evidence but without random sampling.

It might be thought that what we really want is for randomization to lead us always to identify genuine causes rather than simply prima facie causes. But this is also too strong a claim. There is nothing special about genuine causes from the standpoint of randomization. Randomization is as effective in identifying prima facie causes once they are hypothesized as in discovering genuine causes once they too are hypothesized. From the standpoint of experimentation and the role of randomization, the formation of hypotheses about prima facie or genuine causes must be a first step, itself independent of randomization. Then the machinery of randomization and well-designed experiments can be helpful in confirming or disconfirming hypotheses.

It is also worth noting that randomization seems particularly important in highly empirical studies of causes. This is true, for example, of much experimentation in the medical and social sciences. In the case of physics it is often the case that the theoretical structure is so thoroughly developed and at the level of the experiments considered, the causal mechanisms seem to be sufficiently detailed and well understood that randomization as an essential component of experimental design is much less necessary. An essential role of randomization is to provide a method for dealing with unnoticed extraneous causal variables. In the case of physics we often feel much more confident that we can control extraneous causal variables than we do in the case of medical or psychological experiments.

A skeptic, however, could certainly question the distinction I am making here. The complex maze of experimental equipment used in most modern physical experiments is enough to generate doubt about proper control of extraneous causal variables. As the saying goes in some physical laboratories, "if it works don't ask any questions but put a blue ribbon around it" in reference to a complex and fragile piece of equipment. A skeptic might also say that we do not sample particles or other bits of matter in conducting physical experiments because we really make use of the principle of indifference. We have no basis for thinking one particle is different from another. This is due to the fact that we do not have the detailed knowledge of particles or most other bits of matter that we do of human beings or other organisms where it is easy for us to distinguish between them on the basis of salient and readily perceived properties. This argument from ignorance as a basis for not being more careful in sampling in physical experiments has not often been made, and I do not intend to pursue it here. I am sure that some interesting historical cases of physical experiments could be collected in support of it.

4. COMPLEXITY

The definition of randomness in terms of high complexity, which originates in the work of Kolmogorov, Martin-Löf, Cheitin and others, suggests that we could replace random design schemes by complex ones. Suppose, for example, that in a medical experiment we have 200 pairs of blocked subjects. Our original design objective is randomly to assign one member to the experimental group. If we use a simple randomization procedure there are 2^{200} possible assignments. One of these possible assignments would place the first member of each pair in the experimental group for the first 50 pairs, not so for the next 50, so for the next 50, and not so for the final 50. Many experimental scientists would be unhappy with these simply described systematic results of a random procedure. My proposal is to change the method of constructing a sample. First we eliminate from consideration those results that, like the example just given, have low complexity. A la Kolmogorov and others, we measure complexity of a sequence by the length of a program in some standard computer language for describing the sequence. I shall not enter into the technical details here, but note that for practical applications of the kind being discussed, complexity measures for large n are not enough. We would need to fix the language in advance on intuitive grounds of reasonableness—e.g., it did not encode in some bizarre and economical way sequences that in most programming languages would have long descriptions. Given the chosen language we could then throw out the 2^{100}, say, sequences with the lowest complexity measure. Since $2^{100} - 1$ is approximately 2^{100}, we would still have approximately 2^{200} sequences. In any case we would then choose randomly from the remaining set of sequences, which we could do by constructing the sample sequence from a table of random numbers, but then throwing out any constructed sequence whose complexity was below the agreed-upon complexity threshold. Robustness is evident here. The exact complexity number chosen as the threshold is not important. Elimination of the 2^{100} simplest sequences violates the initial distribution assumptions, but the violation leaves us with a reasonable approximation, and it is not worth the effort to compute with the new more complex distribution.

The tension between randomness and complexity is easy to identify. The sampling *procedure* is random. Any sequence is as likely as any other, simple or complex. But the *result* of using the procedure is a given sequence whose complexity can be measured. In the standard informal statistical usage, randomness is a property of certain procedures, complexity a measure of the result.

It may be desirable to modify random sampling procedures in the manner indicated to guarantee complexity of the result, which theoretically is guaranteed in the limit, but not for fixed finite samples. The virtues of random sampling argued for earlier remain if we restrict ourselves to complex results.

17

PROPENSITY REPRESENTATIONS OF PROBABILITY

In recent years a propensity interpretation of probability, thought of primarily as an objective interpretation, has become popular with a number of philosophers, and there has developed a rather large philosophical literature on it. The concept of propensity itself is not a new one. The *Oxford English Dictionary* cites clear and simple uses of *propensity* and its general meaning already in the seventeenth century, for example, from 1660 'Why have those plants... a propensity of sending forth roots?' So the idea of this interpretation of probability is to use the general physical idea of objects having various physical propensities, for example, a propensity to dissolve when placed in water, and to extend this idea to that of probability. As is also clear from these discussions, propensities can be looked upon as dispositions (a rather detailed discussion of this point can be found in Chapter 4 of Mellor, 1971).

The most prominent advocate of the propensity interpretation of probability has been Popper, who set forth the main ideas in two influential articles (1957, 1959). Popper gives as one of his main motivations for developing a propensity interpretation the need to give objective interpretation of single-case probabilities in quantum mechanics, that is, an

*Reprinted from *Erkenntnis*, **26** (1987), 335-358.

objective interpretation of the probabilities of individual events. Single-case probabilities, as they are called in the literature, are of course no problem for subjectivists, but they have been a torturous problem for relative-frequency theorists. A good detailed discussion of how we are to think of propensities in relation both to singular-case probabilities and to relative frequencies is to be found in Giere (1973). I agree with Giere that one of the real reasons to move away from the relative-frequency theory is the single-case problem and therefore we should regard as fundamental or primary the propensity interpretation of singular events. Giere gives a number of textual quotations to show that Popper wavers on this point.

As I pointed out in Suppes (1974c), what is missing in these excellent intuitive discussions of the philosophical and scientific foundation of a propensity interpretation is any sense that there needs to be something proved about the propensity interpretation. Within the framework I propose, this would amount to proving a representation theorem. In Popper's (1974) response to my criticisms he mentions his own ideas about conditional probability and how to handle the problem of evidence that has zero probability. There are some other fine points as well, but the real point of my criticism he misses in not giving a conceptually different analysis of propensities, so that something of genuine interest can be proved about the interpretation. The request for such a proof is not an idle or merely formal request. In order for an interpretation of probability to be interesting, some clear concepts need to be added beyond those in the formal theory as axiomatized by Kolmogorov. The mere hortatory remark that we can interpret propensities as probabilities directly, which seems to be a strong propensity of some advocates, is to miss the point of giving a more thorough analysis.

Because I think there are many good things about the propensity interpretation, as I indicated in my 1974 article on Popper, I want to prove three different representation theorems, each of which is intended to give an analysis for propensity that goes beyond the formal theory of probability, and to include as well, as a fourth example, a surprising theorem from classical mechanics.

The first representation theorem is closest to probability itself and is for radioactive phenomena. The second is for psychological phenomena where propensity is represented in terms of strength of response. The probabilities are then derived explicitly from response strengths. The third example, the most important historically, is the derivation of the behavior of coins, roulette wheels and similar devices, purely on the basis of considerations that belong to classical mechanics. The fourth example shows how random phenomena can be produced by purely deterministic systems.

These four different examples of propensity representation theorems do not in any direct sense force the issue between single-case and relative-frequency views of propensity. But I certainly see no general reason for not using them to compute single-case probabilities. It seems natural to do so whenever a relatively detailed account of the structure of a given propensity is given.

1. PROPENSITY TO DECAY

Before turning to technical developments, there are some general remarks to be made about the approach followed here, which are largely taken from Suppes (1973b).

The first remark concerns the fact that in the axioms that follow, propensities as a means of expressing qualitative probabilities are properties of events and not of objects. Thus, for example, the primitive notation is interpreted as asserting that the event A, given the occurrence of the event B, has a propensity to occur at least as great as the event C, given the occurrence of the event D. Moreover, the events B and D may not actually occur. What we are estimating is a tendency to occur or a propensity to occur, given the occurrence of some other event. If the use of the language of propensity seems awkward or unnatural in talking about the occurrence of events it is easy enough simply to use qualitative probability language and to reserve the language of propensity for talking about the properties of objects, although I am opposed to this move myself. In any case, the issue is peripheral to the main problem addressed. The second remark concerns the clear distinction between the kind of representation theorem obtained here and the sort of theorem ordinarily proved for subjective theories. It is characteristic of subjective theories to prove that the structural axioms impose a unique probability measure on events. It is this uniqueness that is missing from the objective theory as formulated here, and in my own judgment this lack of uniqueness is a strength and not a weakness. Take the case of radioactive decay. From the probabilistic axioms without specific experimentation or identification of the physical substance that is decaying, we certainly do not anticipate being able to derive *a priori* the single parameter of the geometric distribution for decay. It is exactly such a parametric result, i.e., uniqueness up to a set of parameters, that is characteristic of objective theory, and, I would claim, characteristic of standard experimentation in broad areas of science ranging from physics to psychology. In other words, the structural axioms of probability, together with the necessary ones, fix the parametric forms of the probability measure but do not determine it uniquely.

Specific experiments and specific methods of estimation of parameters on the basis of empirical data are needed to determine the numerical values of the parameters. Subjectivists often end up with a working procedure similar to the present one by assuming a probability distribution over the space of parameters. Such procedures are close to what is being discussed here, and well they might be, because there is great agreement on how one proceeds in practice to estimate something like a parameter of a geometric distribution. The important point I want to emphasize is that, in the fundamental underpinnings of the theory, subjectivists have ordinarily insisted upon a unique probability measure, and this commitment to an underlying unique probability measure seems to me to be an unrealistic premise for most scientific applications. It is not that there is any inconsistency in the subjectivistic approach; it is simply that the present objectivistic viewpoint is a more natural one from the standpoint of ordinary scientific practice.

I emphasize also that the intuitive meaning of the weaker structural axioms of objective theory is different from that of the stronger axioms of subjective theory. The objective structural axioms are used to express specific qualitative hypotheses or laws about empirical phenomena. Their form varies from one kind of application to another. A specific structural axiom provides a means of sharply focusing on what fundamental empirical idea is being applied in a given class of experiments. More is said about this point later for the particular case of radioactive decay.

I turn now to the formal developments. First of all, in stating the necessary axioms I shall use qualitative probability language of the sort that is standard in subjective theories of probability. The real place for propensity comes in the discussion of the rather particular structural axioms which reflect strong physical hypotheses about the phenomena of radioactive decay.

The axioms given in the following definition will not be discussed because they are of a type quite familiar in the literature. The ones I give here represent only a minor modification of the first six axioms of Definition 8 of Krantz et al. (1971, p. 222), plus an alternative axiom that is Axiom 7 below. Note that $\approx$ is equivalence, i.e.,

$$A|B \approx C|D \text{ iff } A|B \succeq C|D \text{ and } C|D \succeq A|B.$$

DEFINITION 1. *A structure* $\mathcal{X} = (X,\mathcal{F},\succeq)$ *is a* qualitative probability structure *if and only if* $\mathcal{F}$ *is a* σ*-algebra of sets on* X *and for every* A, B, C, D, E, F, G, A_i, $B_i, i = 1, 2, \ldots$, *in* $\mathcal{F}$ *with* $B, D, F, G \succ \emptyset$ *the following axioms hold*:

Axiom 1. *If* $A|B \succeq C|D$ *and* $C|D \succeq E|F$ *then* $A|B \succeq E|F$;

Axiom 2. $A|B \succeq C|D$ *or* $C|D \succeq A|B$;

Axiom 3. $X \succ \emptyset$;

Axiom 4. $X|B \succeq A|D$;

Axiom 5. $A \cap B|B \approx A|B$;

Axiom 6. *If* $A_i \cap A_j = B_i \cap B_j = \emptyset$ *for* $i \neq j$ *and* $A_i|D \succeq B_i|F$ *for* $i = 1, 2, \ldots,$ *then* $\cup A_i|D \succeq \cup B_i|F$; *moreover, if for some* i, $A_i|D \succ B_i|F$, *then* $\cup A_i|D \succ \cup B_i|F$;

Axiom 7. *If* $A \subseteq B \subseteq D$, $E \subseteq F \subseteq G$, $A|B \succeq E|F$ *and* $B|D \succeq F|G$ *then* $A|D \succeq E|G$; *moreover,* $A|D \succ E|G$ *unless* $A \approx \emptyset$ *or both* $A|B \approx E|F$ *and* $B|D \approx F|G$;

Axiom 8. *If* $A \subseteq B \subseteq D$, $E \subseteq F \subseteq G$, $B|D \succeq E|F$ *and* $A|B \succeq F|G$, *then* $A|D \succeq E|G$; *moreover, if either hypothesis is* $\succ$, *then the conclusion is* $\succ$.

To say that the axioms of Definition 1 are "necessary" means, of course, that they are a mathematical consequence of the assumption that a standard probability measure P is defined on $\mathcal{F}$ such that

$$A|B \succeq C|D \text{ iff } P(A|B) \geq P(C|D). \tag{1}$$

Precisely which necessary axioms are needed in conjunction with the sufficient structural axioms to guarantee the existence of a probability measure satisfying (1) will vary from case to case. It is likely that most of the eight axioms of Definition 1 will ordinarily be needed. In many cases an Archimedean axiom will also be required; the formulation of this axiom in one of several forms is familiar in the literature. The following version is taken from Krantz et al. (1971, p. 223).

DEFINITION 2. *A qualitative conditional probability structure* $\mathcal{X} = (X, \mathcal{F}, \succeq)$ *is* Archimedean *if and only if every standard sequence is finite, where* $(A_1, A_2, \ldots)$ *is a standard sequence iff for all* $i, A_i \succ \emptyset, A_i \subseteq A_{i+1}$ *and* $X|X \succ A_i|A_{i+1} \approx A_1|A_2$.

I turn now to a theorem that seems to have been first stated in the literature of qualitative probability in Suppes (1974c). The reason it did not appear earlier seems to be that the solvability axioms ordinarily used in subjective theories of probability often do not hold in particular physical situations when the probabilistic considerations are restricted just to that situation. Consequently the familiar methods of proof of the existence of a representing probability measure must be changed. Without this change, the theorem is not needed.

The theorem is about standard sequences. Alone, the theorem does not yield a probability measure, but it guarantees the existence of a numerical function that can be extended to all events, and thereby it becomes a measure when the physical hypotheses expressing structural, nonnecessary constraints are sufficiently strong.

THEOREM 1. (*Representation Theorem for Standard Sequences*). *Let* $(A_1, \ldots, A_n)$ *be a finite standard sequence, i.e.,* $A_i \succ \emptyset, A_i \subseteq A_{i+1}$, *and* $X|X \succ A_i|A_{i+1} \approx A_1|A_2$. *Then there is a function* W *such that*

(i) $A_i \subseteq A_j$ *iff* $W(A_i) \leq W(A_j)$,

(ii) *if* $A_i \subseteq A_j$ *and* $A_k \subseteq A_1$ *then*

$$A_i|A_j \approx A_k|A_1 \textit{ iff } W(A_i)/W(A_j) = W(A_k)/W(A_1).$$

Moreover, for any W *satisfying (i) and (ii) there is a* q *with* $0 < q < 1$ *and a* $c > 0$ *such that*

$$W(A_i) = cq^{n+1-i}.$$

Proof. Let $0 < q < 1$. Define $W(A_i)$ as

$$W(A_i) = q^{n+1-i}. \tag{1}$$

Then obviously (i) is satisfied, since the members of a standard sequence are distinct; otherwise there would be an i such that $A_i = A_{i+1}$ and thus $A_i|A_{i+1} \approx X|X$, contrary to hypothesis. So we may turn at once to (ii). First, note the following.

$$A_i|A_{i+m} \approx A_j|A_{j+m}. \tag{2}$$

The proof of (2) is by induction. For $m = 1$, it is just the hypothesis that for every $i, A_i|A_{i+1} \approx A_1|A_2$. Assume now it holds for $m - 1$; we then have

$$A_i|A_{i+(m-1)} \approx A_{j+(m-1)},$$

and also for any standard sequence

$$A_{i+(m-1)}|A_{i+m} \approx A_{j+(m-1)}|A_{j+m},$$

whence by Axiom 7, $A_i|A_{i+m} \approx A_j|A_{j+m}$, as desired. Next, we show that if $A_i \subseteq A_j, A_k \subseteq A_l$ and $A_i|A_j \approx A_k|A_l$, then there is an $m \geq 0$ such that $j = i + m$ and $l = k + m$. Since $A_i \subseteq A_j$ and $A_k \subseteq A_l$, there must be nonnegative integers m and m' such that $j = i + m$ and $l = k + m'$.

Suppose $m \neq m'$, and without loss of generality suppose $m + h = m'$, with $h > 0$. Then obviously

$$A_i | A_{i+m} \approx A_k | A_{k+m}.$$

In addition,

$$A_{i+m} | A_{i+m} \approx X | X \succ A_{k+m} | A_{k+m+h},$$

and so again by Axiom 7

$$A_i | A_{i+m} \succ A_k | A_{k+m'},$$

contrary to our hypothesis, and so we must have $m = m'$.

With these results we can establish (ii). We have as a condition that $A_i \subseteq A_j$ and $A_k \subseteq A_l$. Assume first that $A_i | A_j \approx A_k | A_l$. Then we know that there is an m such that $j = i + m$ and $l = k + m$, whence

$$\begin{aligned} W(A_i)/W(A_j) &= q^{n+1-i}/q^{n+1-i-m} \\ &= q^{n+1-k}/q^{n+1-k-m} \\ &= W(A_k)/W(A_l). \end{aligned}$$

Second, we assume that

$$W(A_i)/W(A_j) = W(A_k)/W(A_l).$$

From the definition of W it is a matter of simple algebra to see that there must be an m' such that $j = i + m'$ and $l = k + m'$, whence by our previous result, $A_i | A_j \approx A_k | A_l$.

Finally, we must prove the uniqueness of q as expressed in the theorem. Suppose there is a W' satisfying (i) and (ii) such that there is no $c > 0$ and no q such that $0 < q < 1$ and for all i

$$W'(A_i) = cq^{n+1-i}.$$

Let

$$\begin{aligned} W'(A_n) &= q_1 \\ W'(A_{n-1}) &= q_2. \end{aligned}$$

Let

$$\begin{aligned} q &= \frac{q_2}{q_1} \\ c &= \frac{q_1^2}{q_2}. \end{aligned}$$

Obviously,

$$\begin{aligned} W'(A_n) &= cq \\ W'(A_{n-l}) &= cq^2. \end{aligned}$$

On our supposition about W', let i be the largest integer (of course $i \preceq n$) such that

$$W'(A_i) \neq cq^{n+1-i}.$$

We have

$$A_i|A_{i+1} \approx A_{n-1}|A_n,$$

whence by (ii)

$$W'(A_i)/cq^{n-i} = cq^2/cq,$$

and so

$$W(A_i) = cq^{n+1-i},$$

contrary to hypothesis, and the theorem is proved.

I turn now to radioactive decay phenomena. One of the best-known physical examples of a probabilistic phenomenon for which no one pretends to have a deeper underlying deterministic theory is that of radioactive decay. Here I shall consider for simplicity a discrete-time version of the theory which leads to a geometric distribution of the probability of decay. Extension to continuous time is straightforward but will not be considered here. In the present context the important point is conceptual, and I want to minimize technical details. Of course, the axioms for decay have radically different interpretations in other domains of science, and some of these will be mentioned later.

In a particular application of probability theory, the first step is to characterize the sample space, i.e., the set of possible experimental outcomes, or as an alternative, the random variables that are numerical-valued functions describing the phenomena at hand. Here I shall use the sample-space approach, but what is said can easily be converted to a random-variable viewpoint.

From a formal standpoint, the sample space X can be taken to be the set of all infinite sequences of 0's, and 1's containing exactly one 1. The single 1 in each sequence occurs as the nth term of the sequence representing the decay of a particle on the nth trial or during the nth time period, with its being understood that every trial or period is of the same duration as every other. Let E_n be, then, the event of decay on trial n. Let W_n be the event of no decay on the first n trials, so that

$$W_n = -\bigcup_{i=1}^{n} E_i.$$

The single structural axiom is embodied in the following definition. The axiom just asserts that the probability of decay on the nth trial, given that decay has not yet occurred, is equivalent to the probability of decay on the first trial. It thus expresses in a simple way a qualitative principle of constancy or invariance of propensity to decay through time.

DEFINITION 3. *Let X be the set of all sequences of 0's and 1's containing exactly one 1, and let $\mathcal{F}$ be the smallest σ-algebra on X which contains the algebra of cylinder sets. A structure $\mathcal{X} = (X,\mathcal{F},\succeq)$ is a* qualitative waiting-time structure with independence of the past *iff $\mathcal{X}$ is a qualitative conditional probability structure and in addition the following axiom is satisfied for every n, provided $W_{n-l} \succ \emptyset$:*

Waiting-time Axiom. $E_n|W_{n-1} \approx E_1$.

The structures characterized by Definition 3 are called *waiting-time structures with independence of the past*, because this descriptive phrase characterizes their general property abstracted from particular applications like that of decay.

The simplicity of this single structural axiom may be contrasted with the rather involved axioms characteristic of subjective theories of probability. In addition, the natural form of the representation theorem is different. The emphasis is on satisfying the structural axioms—in this case, the waiting-time axiom—and having a unique parametric form, rather than a unique distribution.

THEOREM 2. *(Representation Theorem for Decay). Let $\mathcal{X} = (X,\mathcal{F},\succeq)$ be a qualitative waiting-time structure with independence of the past. Then there exists a probability measure on $\mathcal{F}$ such that the waiting-time axiom is satisfied, i.e.,*

$$\text{(i)} \qquad P(E_n|W_{n-1}) = P(E_i),$$

and there is a number p with $0 < p \leq 1$ such that

$$\text{(ii)} \qquad P(E_n) = p(1-p)^{n-1}.$$

Moreover, any probability measure satisfying (i) is of the form (ii).

Proof. The events E_n uniquely determine an atom or possible experimental outcome x of X, i.e., for each n, there is an x in X such that

$$E_n = \{x\},$$

a situation which is quite unusual in sample spaces made up of infinite sequences, for usually the probability of any x is strictly zero.

If $E_1 \approx X$, then $P(E_1) = 1$, and the proof is trivial. On the other hand, if $X \succ E_1$, then for each $n, (W_n, \ldots, W_1)$ is a standard sequence satisfying the hypotheses of the representation theorem for standard sequences. The numbering is inverted, i.e., $W_{i+1} \subseteq W_i$ and $W_{i+1}|W_i \approx$

$W_2|W_1$. (If $(W_1, \ldots, W_n)$ were a standard sequence, then so would be the infinite sequence $(W_1, \ldots, W_n, \ldots)$ in violation of the necessary Archimedean axiom.) That $W_{i+1} \subseteq W_i$ is obvious from the definition of W_i. By virtue of the waiting-time axiom

$$W_{i+1}|W_i \approx W_1,$$

since $E_i = W_{i-1} - W_i$, and $W_1 = -E_1$, so

$$\begin{aligned} E_1 &\approx E_{i+1}|W_i \approx W_i - W_{i+1}|W_i \\ &\approx -W_{i+1}|W_i, \end{aligned}$$

and so by elementary manipulations

$$W_1 \approx -E_1 \approx W_{i+1}|W_i.$$

Using the representation theorem for standard sequences, we know there is a numerical function P' and numbers c and q with $0 < q < 1$ and $c > 0$ such that

$$P'(W_i) = cq^i.$$

(Let i' be the numbering in the reverse order $(n, \ldots, 1)$; then $i' = n - (i - 1)$, and the exponent $n + 1 - 1$ in the representation theorem becomes i'.) Starting with a fixed standard sequence of length n, P' can be extended to every $i > n$ in an obvious fashion.

The next step is to extend P' as an additive set function to all atoms E_i by the equations

$$\begin{aligned} P'(E_i) &= P'(W_{i-1} - W_i) \\ &= P'(W_{i-1}) - P(W_i) \\ &= c(q^{i-1} - q^i) \\ &= cq^{i-1}(1 - q). \end{aligned}$$

The consistency and uniqueness of this extension is easy to check. Now

$$\sum_{i=1}^{\infty} P'(E_i) = c,$$

so we set $c = 1$ to obtain a measure P normed on 1 from P' and let $p = 1 - q$. We then have

$$\begin{aligned} P(E_i) &= p(1-p)^{i-1} \\ P(W_i) &= (1-p)^i. \end{aligned}$$

The function P just defined is equivalent to a discrete density on X, since $E_i = \{x\}$ for some x in X, and thus P may be uniquely extended to the σ-algebra of cylinder sets of X by well-known methods.

Finally, the proof of (ii) is immediate. If we suppose there is an n such that $P(E_n) \neq p(1-p)^{n-1}$, where $p = P(E_1)$, we may use the waiting-time axiom to obtain a contradiction, for by hypothesis $P(W_{n-1}) = (1-p)^{n-1}$ and $P(E_n|W_{n-1}) = p$, whence $P(E_n) = pP(W_{n-1}) = p(1-p)^{n-1}$.

I quite accept that a criticism of this particular representation theorem is that the analysis of propensity is too close to the analysis of probability from a formal standpoint, a complaint I made earlier about some of the literature on propensity. In general, propensities are not probabilities, but provide the ingredients out of which probabilities are constructed. I think the favorable positive argument for what I have done has got to be put on a more subtle and therefore more fragile basis. The point has been made, but I will make it again to emphasize how propensities enter. The waiting-time axiom is a structural axiom that would never be encountered in the standard theory of subjective probability as a fundamental axiom. It is an axiom special to certain physical phenomena. It represents, therefore, a qualitative expression of a propensity. Second, the probabilities we obtain from the representation theorem are not unique but are only unique up to fixing the decay parameter. Again, this is not a subjective concept but very much an objective one. Identifying and locating the number of physical parameters to be determined is a way of emphasizing that propensities have entered and that a purely probabilistic theory with a unique measure has not been given.

2. PROPENSITY TO RESPOND

There is a long history of various theoretical models being proposed in psychology to represent response strength, which in turn is the basis for choice probabilities. By a "choice probability" I mean the probability that a given item a will be selected from a set A in some given experimental or naturalistic setting. The fact that the choice is to be represented by a probability is a reflection that the standard algebraic model of expected utility does not adequately represent much actual behavior. Whatever one may think individuals should do, it is a fact of life, documented extensively both experimentally and in other settings, that individuals, when presented with what appears to be repetitions of the same set of alternatives to choose from, do not repeatedly choose the same thing. The formal study of such situations has also been a matter of intensive work in the past several decades. One of the most simple and elegant models is the choice model proposed by Luce (1959). In Luce's own development he proceeds from his choice axiom, which is stated in terms of observable probabilities, to the existence of response strengths. To illustrate the kind of idea we want here we shall begin the other way, that is, by postulating a response strength, and then showing how from this postulate we easily derive his choice axiom. Second, an advantage of response strengths over response probabilities is to be seen in the formulation of Luce's alpha model of learning, where the transformations that represent learning from one trial to another are linear when formulated in terms of response strengths, but nonlinear when formulated in terms of choice probabilities.

We turn now to the formal development of these ideas. In the intended interpretation T is a presented set of alternatives to choose from, and the numerical function v is the measure of response (or stimulus) strength.

DEFINITION 4. *Let T be a nonempty set, and let v be a nonnegative real-valued function defined on T such that for at least one x in T, $v(x) > 0$, and $\sum_{x \in T} v(x)$ is finite. Then $\mathcal{T}$ $=(T, v)$ is a* response-strength model (of choice).

The general requirements on a response-strength model are obviously very weak, but we can already prove a representation theorem that is more special than earlier ones, in the sense that in addition to satisfying the axioms of finitely additive probability spaces, Luce's Choice Axiom is satisfied as well. Moreover, to get various learning models, we impose further conditions.

THEOREM 3 (*Representation Theorem*). *Let $\mathcal{T} = (T, v)$ be a response-strength model, and define for U in $\mathcal{P}(T)$, the power set of T,*

$$P_T(U) = \sum_{x \in U} v(x) \Bigg/ \sum_{x \in T} v(x).$$

Then $(T, \mathcal{P}(T), P_T)$ is a finitely additive probability space. Moreover, the probability measure P_T satisfies Luce's choice axiom, i.e., for V in $\mathcal{P}(T)$, with $\sum_{x \in V} v(x) \neq 0$, and with v' being v restricted to V, $\mathcal{V} = (V, v')$ is a response-strength model such that for $U \subseteq V$

$$P_V(U) = P_T(U|V).$$

Proof. The general representation part of the theorem is obvious. To prove Luce's axiom, we note that because $U \subseteq V$

$$\begin{aligned} P_T(U|V) &= P_T(U \cap V)/P_T(V) \\ &= \frac{\sum_{x \in U} v(x)}{\sum_{x \in T} v(x)} \Bigg/ \frac{\sum_{x \in V} v(x)}{\sum_{x \in T} v(x)} \\ &= \sum_{x \in U} v(x) \ \Big/ \ \sum_{x \in V} v(x) \\ &= P_V(U). \end{aligned}$$

The notation used in the theorem is slightly subtle and can be initially confusing. Note that the set referred to by the subscript represents the full physical set of choice alternatives. The set conditioned on, V in the case of $P_T(U|V)$, is information about the choice actually occurring from among the elements of a subset of T. It is not at all tautological that Luce's axiom should hold. In fact, there is a sizable statistical literature on how to best estimate response strengths for observed choice probabilities (Bradley and Terry, 1952; Bradley, 1954a, b, 1955; Abelson and Bradley, 1954; Ford, 1957).

To illustrate how the formulation of theory in terms of a propensity can simplify some analyses, we sketch the situation for Luce's alpha model of learning. Let f be the learning function mapping the finite vector of response strengths $v = (v_1, \ldots, v_r)$ from one trial to the next. Here we assume T is finite—in particular has cardinality r. We assume that response strengths are *unbounded*, i.e., for any real number α there is an n such $|f^n(v)| > \alpha$, where f^n represents n iterations of the learning function. Secondly, *superposition* of learning holds, i.e., for any $v, v^* > 0$

$$f(v + v^*) = f(v) + f(v^*).$$

Third, *independence of scale or units* holds, i.e., for $v > 0$ and any real number $k > 0$

$$f(kv) = kf(v).$$

But it is a well-known result in linear algebra that the assumed conditions imply that f is a linear operator on the given r-dimensional vector space. In contrast, under these assumptions but no stronger ones the behavior from trial-to-trial of the response probabilities $P_T(U)$ is complicated; in particular, they are not related by a linear transformation of the probabilities.

Although the proof of Theorem 3 is very simple and in the development thus far little structure has been imposed on the response-strength function, the intended interpretation fits in very nicely with the propensity concept of probability—at least as I envisage the development of representation theorems for various propensities. In fact, a pluralistic aspect of propensities that I like is that there is no single natural representation theorem. Many different physical and psychological propensities should produce unique representation theorems. On the other hand, an obvious deficiency of Theorem 4, and other similarly "direct" representations of probability in terms of some propensity, is that no guarantee of randomness is provided. This is also a deficiency of the radioactive decay example as well. In both cases, adding axioms to deal with randomness is difficult. In the decay case, what is naturally a real-time phenomenon has to be dealt with, rather than the typical multinomial case. In the response-strength models, one would immediately expect learning from repetition and thus the obvious sequences of trials would not represent stationary processes. In principle the standard machinery for defining finite random sequences could be used, but the technical problems of correct formulation seem too numerous to try to solve here.

3. PROPENSITY FOR HEADS

There is a tradition that begins at least with Poincaré (1912) of analyzing physical systems that we ordinarily consider chance devices as classical mechanical systems. More detailed applications to chance devices were given by Smoluchowski (1918) and, in particular, by Hopf (1934). The history of these ideas has been nicely chronicled by von Plato (1983). The simple approach developed here, which requires only Riemann integration, is mainly due to Keller (1986).

We shall analyze the mechanics of coin tossing, but, as might be expected, under several simplifying assumptions that would be only partly satisfied in practice. First, we consider a circular coin of radius a whose thickness is negligible. Second, we assume perfect uniformity in the distribution of the metal in the coin, so that its center of gravity is also its geometrical center. The different marking for a head and a tail is thus assumed to be negligible. Third, we neglect any friction arising from its spinning or falling. Fourth, we carry the analysis only to the first point of impact with the surface on which it lands. We assume it does not change the face up from this point on. We thereby ignore any problems of elasticity that might lead to bouncing off the surface, spinning and landing again before coming to rest. As Hopf points out, real systems are dissipative rather than conservative, but the mathematical analysis that replaces all of the above assumptions with the most realistic ones we can formulate is still not available in complete form. On the other hand, the idealizations we make are not totally unrealistic; there are good physical reasons for believing that more realistic assumptions about dissipation due to friction, etc., would not affect at all the conceptual character of the analysis, but only the quantitative details, which are not critical for our purposes.

It is also useful to ask whether the analysis to be given fits into the standard theory of particle mechanics. The answer is 'almost but not quite'. To take account of spin we treat the coin as a rigid body, not as a particle, although we could imitate the spin properties exactly by a finite collection of particles whose mutual distances remain constant.

Now to the formal details. We use a Cartesian coordinate system with x and z in the horizontal plane and with y being the measure of height, so that $y(t)$ is the height of the center of gravity of the coin at time t. The only vertical force is the force of gravity, so the Newtonian equation of motion is

$$\frac{d^2y(t)}{dt^2} = -g, \tag{1}$$

where g is the constant acceleration of gravity. As initial conditions at time $t = 0$, we suppose the height is a and the toss gives the coin an upward velocity u, i.e.,

$$y(0) = a, \quad \dot{y}(0) = u. \tag{2}$$

Equations (1) and (2) uniquely determine $y(t)$ up to the point of impact. As is easily shown

$$y(t) = -\frac{gt^2}{2} + ut + a. \tag{3}$$

As for spin, we assume the coin is rotating about a horizontal axis which lies along a diameter of the coin; we fix the z-coordinate axis to be parallel to this rotation axis, and we measure angular position as follows. The angle $\emptyset(t)$ is the angle between the positive y-axis and a line perpendicular to the head-side of the coin—both these lines lie in the x-y plane as can be seen from Figure 1, taken from Keller (1986). We assume that initially the coin is horizontal with heads up, and the toss gives it positive angular velocity ω. So that at $t = 0$, the initial spin conditions are:

$$\theta(0) = 0, \quad \dot{\theta}(0) = \omega. \tag{4}$$

Moreover, assuming no dissipation, as we do, the equation governing the rotational motion of the coin is just that of constant velocity.

$$\frac{d^2\theta(t)}{dt^2} = 0. \tag{5}$$

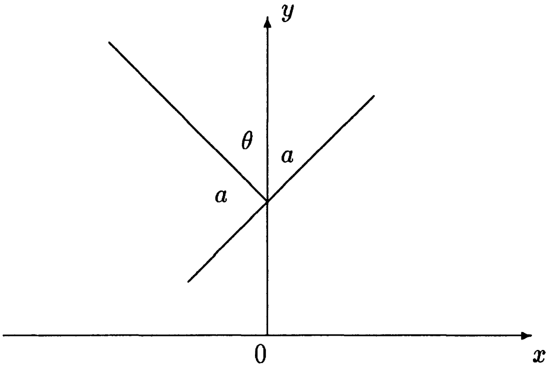

Figure 1. The x, y plane intersects the coin along a diameter of length $2a$. The normal to the side of the coin marked heads makes the angle θ with the positive y-axis.

The unique solution of (4) and (5) is:

$$\theta(t) = \omega t. \tag{6}$$

Let t_s be the point in time at which the coin makes contact with the surface on which it lands, which surface we take to be the plane $y = 0$. Given the earlier stated assumption that the coin does not bounce at all, the coin will have a head up iff

$$2n\pi - \pi/2 < \theta(t_s) < 2n\pi + \pi/2, \quad n = 0, 1, 2, \ldots \tag{7}$$

We now want to find t_s. First, we note that at any time t, the lowest point of the coin is at $y(t) - a|\sin\theta(t)|$. So t_s is the smallest positive root of the equation

$$y(t_s) - a|\sin\theta(t_s)| = 0. \tag{8}$$

We next want to find what Keller calls the *pre-image of heads*, i.e., to find the set of initial values of velocity u and angular velocity ω for which the coin ends with a head up. Let us call H this set of points in the u, ω-plane.

We look first at the endpoints defined by (7). These together with (6) yield – just for the endpoints,

$$\omega t_s = (2n + \tfrac{1}{2}\pi), \tag{9}$$

and also at these endpoints $\sin\theta(t_s) = \pm 1$, so (8) in these cases simplifies to

$$y(t_s) - a = 0, \tag{10}$$

which combined with (5) yields

$$ut_s - gt_s^2/2 = 0. \tag{11}$$

The importance of the endpoints is that they determine the boundaries of the region H. In particular, we examine solutions of (11) to determine t_s and then (9) to determine the boundaries. Equation (11) has two solutions:

$$t_s = 0, \quad t_s = 2u/g.$$

The first one yields only trivial results, so we use the second solution in (9) to obtain a relation in terms only of ω and u:

$$\omega = \frac{(2n \pm \frac{1}{2})\pi g}{2u}, \quad n = 0, 1, 2, ... \tag{12}$$

The relationship (12) is graphed in Figure 2 (after Keller, 1986) for various values of n. As can be seen from (12), each curve is a hyperbola. On the axis $\omega = 0$, a head remains up throughout the toss, so the strip bordered by the axis belongs to H. The next strip is part of T, the complement of H, and as is obvious the alternation of strips being either part of H or T continues. From (12) we can infer that the strips are of equal vertical separation, namely, $\pi g/2u$, except for $n = 0$ for the lowest one where the vertical distance from the axis is $\pi g/4u$.

The critical observation is that, as the initial velocity u of the toss increases, the vertical separation decreases and tends to zero. This means that the alternation between H and T is generated by small changes in u.

As we shall show, essentially any mathematically acceptable probability distribution of u and ω, at time $t = 0$, will lead to the probability of heads being 0.5. The mechanical process of flipping dominates the outcome. Small initial variations in u and ω, which are completely unavoidable, lead to the standard chance outcome. Put explicitly, the standard mechanical process of flipping a coin has a strong propensity to produce a head as outcome with probability 0.5.

To calculate P_H, the probability of heads, we assume an initial continuous probability distribution $p(u,\omega)$, with, of course, $u > 0$ and $\omega > 0$. It is important to note that no features of symmetry of any kind are assumed for $p(u,\omega)$. We have at once

$$P_H = \int\int_H p(u,\omega)d\omega\ du. \tag{13}$$

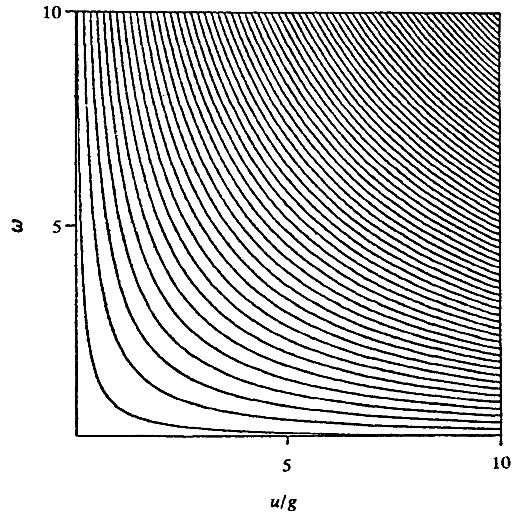

Figure 2. The curves which separate the sets H and T, the pre-images of heads and tails in the u, ω plane of initial conditions, are shown for various values of n, with the abscissa being u/g.

It is obvious that (13) imposes no restriction on P_H. Given any value of P_H desired, i.e., any probability of heads, we can find a distribution of $p(u, \omega)$ that will produce it in accordance with (13).

What can be proved is that as the velocity u increases we can prove that in the limit $P_H = \frac{1}{2}$. The actual rate of convergence will be sensitive to the given distribution $p(u, \omega)$.

THEOREM 4 (*Representation Theorem*).

$$\lim_{U \to \infty} P(H|u > U) = \frac{1}{2}.$$

Proof. We first write the conditional probability without taking the limit:

$$(14) \qquad P(H|u > U) = \frac{\displaystyle\int_U^\infty \sum_{n=0}^{\infty} \int_{(2n-(1/2))\pi g/2u}^{(2n+(1/2))\pi g/2u} p(u,\omega) d\omega du}{\displaystyle\int_U^\infty \int_0^\infty p(u,\omega) d\omega du}$$

(What has been done in the numerator is to integrate over each "slice" of H, given by (12).) The set T is the complement of H, and so the boundaries of slices of T are:

$$\psi = \frac{(2n+1 \pm \frac{1}{2})\pi g}{2u}, \tag{15}$$

and we can write the denominator of (14) as:

$$\begin{aligned} &\int_U^\infty \sum_{n=0}^\infty \int_{(2n-(1/2))\pi g/2u}^{(2n+(1/2))\pi g/2u} p(u,\omega)d\omega \; du + \int_U^\infty \sum_{n=0}^\infty \\ &\quad \times \int_{(2n+(1/2))\pi g/2u}^{(2n+(3/2))\pi g/2u} p(u,\omega)d\omega \; du. \end{aligned} \tag{16}$$

We next want to evaluate the numerator of the right-hand side of (14) as $U \to \infty$. From the definition of the Riemann integral, we can approximate each integral inside the summation side by the length of the interval, which is $\Pi g/2u$, and the value of the integral at the midpoint of the interval, which is $p(u, n\pi g/u)$.

$$\begin{aligned} &\lim\nolimits_{U\to\infty} \int_U^\infty \sum_{n=0}^\infty \int_{(2n-(1/2))\pi g/2u}^{(2n+(1/2))\pi g/2u} p(u,\omega)d\omega \; du \\ &= \lim\nolimits_{U\to\infty} \int_U^\infty \sum_{n=0}^\infty p\left(u, \frac{n\pi g}{u}\right) \frac{\pi g}{2u} du. \end{aligned} \tag{17}$$

And it is straightforward to show that as $U \to \infty$ the integral on the right converges to $\frac{1}{2}P(u > U)$. From this result, (14), (16), and (17), we have as desired:

$$\lim_{U\to\infty} P(H|u > U) = \frac{\lim_{U\to\infty} \frac{1}{2}P(u > U)}{\lim_{U\to\infty} \frac{1}{2}P(u > U) + \lim_{U\to\infty} \frac{1}{2}P(u > U)} = \frac{1}{2}.$$

4. PROPENSITY FOR RANDOMNESS

This fourth example is a special case of the three-body problem, certainly the most extensively studied problem in the history of mechanics. Our special case is this. There are two particles of equal mass m_1 and m_2 moving according to Newton's inverse-square law of gravitation in an elliptic orbit relative to their common center of mass which is at rest. The third particle has a nearly negligible mass, so it does not affect the motion of the other two particles, but they affect its motion. This third particle is moving along a line perpendicular to the plane of motion of the first two particles and intersecting the plane at the center of their mass—let this be the z axis. From symmetry considerations, we can see that the third particle will not move off the line. The restricted problem is to describe the motion of the third particle.

To obtain a differential equation in simple form, we normalize the unit of time so that the temporal period of rotation of the two masses in the x, y-plane is 2π, we take the unit of length to be such that the gravitational constant is one, and finally $m_1 = m_2 = \frac{1}{2}$, so that $m_1 + m_2 = 1$. The force on particle m_3, the particle of interest, from the mass of particle 1 is:

$$F_1 = -\frac{m_1}{z^2 + r^2} \cdot \frac{(z, r)}{\sqrt{z^2 + r^2}},$$

where r is the distance in the x, y-plane of particle 1 from the center of mass of the two-particle system m_1 and m_2, and this center is, of course, just the point $z = 0$ in the x, y plane. Note that $(z, r)/\sqrt{z^2 + r^2}$ is the unit vector of direction of the force F_1. Similarly,

$$F_2 = -\frac{m_2}{z^2 + r^2} \cdot \frac{(z, -r)}{\sqrt{z^2 + r^2}}.$$

So, simplifying, we obtain as the ordinary differential equation of the third particle

$$\frac{d^2 z}{dt^2} = -\frac{z}{(z^2 + r^2)^{3/2}}.$$

The analysis of this easily described situation is quite complicated and technical, but some of the results are simple to state in informal terms. Near the escape velocity for the third particle—the velocity at which it leaves and does not periodically return, the periodic motion is very irregular. In particular, the following remarkable theorem can be proved. Let $t_1, t_2, \ldots$ be the times at which the particle intersects the plane of motion of the other two particles. Let s_k be the largest integer equal to or less than the difference between t_{k+1} and t_k times a constant.[1] Variation in the s_k's obviously measures the irregularity in the periodic motion. The theorem, due to the Russian mathematicians Sitnikov (1960) and Alekseev (1969a, b), as formulated in Moser (1973), is this.

THEOREM 5. *Given that the eccentricity of the elliptic orbits is positive but not too large, there exists an integer, say α, such that any infinite*

[1] The constant is the reciprocal of the period of the motion of the two particles in the plane.

sequence of terms s_k with $s_k \geq \alpha$, corresponds to a solution of the deterministic differential equation governing the motion of the third particle.[2]

A corollary about random sequences immediately follows. Let s be any random sequence of heads and tails—for this purpose we can use any of the several variant definitions—Church, Kolmogorov, Martin-Löf, etc. We pick two integers greater than α to represent the random sequence—the lesser of the two representing heads, say, and the other tails. We then have:

COROLLARY. *Any random sequence of heads and tails corresponds to a solution of the deterministic differential equation governing the motion of the third particle.*

In other words, for each random sequence there exists a set of initial conditions that determines the corresponding solution. Notice that in essential ways the motion of the particle is completely unpredictable even though deterministic. This is a consequence at once of the associated sequence being random. It is important to notice the difference from the earlier coin-flipping case, for no distribution over initial conditions and thus no uncertainty about them is present in this three-body problem. No single trajectory in the coin-flipping case exhibits in itself such random behavior.

In this fourth case, propensity may be expressed as the tendency of the third particle to behave with great irregularity just below its escape velocity. In ordinary terms, we might want to say that the propensity is one for irregularity, but of course we can say more for certain solutions, namely, certain initial conditions lead to a propensity to produce random behavior.

[2] The correspondence between a solution of the differential equation and a sequence of integers is the source of the term *symbolic dynamics*. The idea of such a correspondence originated with G. D. Birkhoff in the 1930s.

18

INDETERMINISM OR INSTABILITY, DOES IT MATTER?

1. SKEPTICISM ABOUT DETERMINISM

In my recent book, *Probabilistic Metaphysics* (1984), I have argued at some length against determinism as a viable philosophical or scientific thesis. I want first to review those arguments and then go on to look at an alternative way of viewing phenomena. Instead of the dichotomy deterministic or indeterministic, perhaps the right one is stable or unstable.

In expressing my skepticism about determinism I shall not linger over a technically precise definition. It seems to me that the intuitive notion that phenomena are deterministic when their past uniquely determines their future will serve quite adequately in the present context.

The natural basis of skepticism is our remarkable inability to predict almost any complete phenomenon of interest, and even more, our inability to write down adequate difference or differential equations. Consider, for

*Reprinted from *Causality, method and modality* (ed. by G. G. Brittan, Jr.), 1991, pp. 5-22. Dordrecht, Netherlands: Kluwer Academic Publishers. The first draft of this paper was read at a symposium on indeterminism at the annual meeting of the Pacific Division of the American Philosophical Association, March 23, 1985. A later draft was the basis of a lecture at the College de France at the invitation of Jules Vuillemin on May 12, 1986. It is a pleasure to dedicate this paper to him. Our many conversations on a variety of philosophical topics have both enlightened and delighted me.

example, a gust of wind and its effect on leaves of grass, the branches of a tree, the particles of dust agitated in various ways. It seems utterly out of the question to predict these effects in any detail. Moreover, it seems hopeless even to think of writing down the equations, let alone solving them. It might be noted the particles of dust, at least, would be within the range of the phenomena of Brownian motion, and the hopelessness of actually predicting such motion has been recognized for a long time. Of course, this example of Brownian motion raises a problem that needs remarking. One standard view of classical physics is that all phenomena are deterministic—we are just unable to analyze some phenomena in adequate detail. But even here there is reason for skepticism. The standard result of the standard theory of Brownian motion is that because of the high incidence of collisions the path of a particle is continuous but differentiable almost nowhere (only on a set of measure zero). Given that the path is this kind of trajectory, it becomes obvious that determinism is out of the question just because of the many collisions. It is a familiar fact of classical mechanics that collisions in general cause great difficulty for deterministic theorems. The kind of result that we have in the case of Brownian motion is not just a matter of difficulty, it is a matter of principled hopelessness. So I take it that insofar as the phenomena I have just described fall within the purview of the theory of Brownian motion, determinism is ruled out.

For many familiar human phenomena we do not even have the elements of schematic analysis given by the probabilistic theory of Brownian motion. Examples are easy to think of. A favorite of mine is the babble of speech. The idea of ever being able to determine the flow of talk even between just one set of persons, not to speak of a billion, given whatever knowledge you might hope to have seems ridiculous and absurd. There is no reason whatsoever to think we will ever have theories that lead to deterministic results. It is certainly true that in occasional high states of deliberation we formulate very carefully the words we are going to utter, but this is not the standard condition of speech. Moreover, even in such states of high deliberation we do not and are not able consciously to control the prosodic contours of the utterance. In fact, as we descend from the abstract talk of grammarians and model theorists concerned with semantics to the intricate details of the actual sound-pressure waves emitted by speakers and received by listeners, the problem of having a deterministic theory of speech looms ever more hopeless.

I have the same skepticism toward deterministic theories of vision. Such a theory for any serious level of detail seems out of the question. The reasons for thinking this are many in number. The long history of theories of vision and the difficulties we still have in giving detailed partial descriptions of what the visual system is sensing provide some evidence. Detailed physiological studies showing that the human eye is sensitive to even a single photon provide other kinds of evidence, as do quantitative studies of eye and head movements. The extraordinarily complicated nature of the transduction that takes place in the optical system in order to send messages to the central nervous system is another case in point. Someone might want to claim that we could have a gross deterministic theory of vision, but such a theory would be superficial and uninteresting. The actual mechanisms seem intrinsically subtle and complex. Of course, there are some kinds of complex problems that we feel confident in tackling, but anyone who has taken a serious look at problems of vision will back away rather rapidly from optimistic claims about having within the framework of contemporary science, or science as we can foresee it to be in the future, a workable, detailed deterministic theory.

What I have had to say about speech and vision applies also to the sense of smell. The evidence seems pretty good that this sense is sensitive down to the presence of a substance at the molecular level. Moreover, what theories there are of the activities of single recognition cells are probabilistic in character. As far as I know, no one has attempted to propose a serious deterministic theory of smell.

These familiar phenomena I am using to buttress my reasons for skepticism about determinism are easily matched by a dozen others. Given the extraordinarily small number of phenomena about which we can have a deterministic phenomena, there is cause for psychological and philosophical speculation as to why the concept of determinism has ever achieved the importance it has in our thinking about the world around us. To adopt a broad deterministic view toward the world does require not quite the extreme faith of the early Christians, but at least that of such diverse eighteenth-century optimists as Kant and Laplace.[1]

Surely one psychological root of the faith in determinism is its conflation with prediction. Hegel (1899, p. 278) reports that Napoleon in a conversation with Goethe remarked that the conceptual role of fate in the ancient world has been replaced by that of bureaucratic policy in modern times—with the implication that uniform predictability of individual behavior subject to the bureaucracy is, in principle, what we can now have. The search for methods of prediction has ranged from zodiacs to chicken gizzards and is found in every land. The primitive urge to know the future has in no way been stilled by modern science, but only rechanneled into more austere forms. The new skepticism, so I am arguing, should be about the omnipresence of determinism, not the omniscience of God.

[1] Historically we probably need to think of Kant as a cryptodeterminist.

2. HOW TO SAVE DETERMINISM

Before making some direct comparisons with indeterminism in terms of instability there are some preliminary points to be made about unstable systems. The intuitive idea of instability in mechanics is this. Wide divergence in the behavior of two systems identical except for initial conditions is observed even when the initial conditions are extremely close. There are two aspects of unstable systems that make prediction of their behavior difficult, and therefore make difficult the realization of the deterministic program, even if the systems are, in fact, deterministic. One source of difficulty is that the initial conditions can be measured only approximately. If a system is not stable in the appropriate sense—I omit a technical definition here but it is straightforward to give one—, it will be impossible to predict its behavior for any but short intervals of time with any accuracy. In this case, we attribute predictive failures to a possibly small uncertainty in the initial conditions. We shall leave aside in the present discussion whether this uncertainty should be treated epistemologically or ontologically. Some later remarks will have something more to say about this issue.

A second aspect of an unstable system can be that the solutions are not given in closed form, and calculations based on various methods of series expansion, etc., will not give accurate predictions. In other words, we cannot count on numerical methods to give us a detailed result for periods of prediction of any length. If the system is unstable, the accumulation of small errors in numerical methods of approximation, which may be the only ones available, can lead to unavoidable problems of accuracy. This last problem is especially true of systems that are governed by nonlinear differential equations.

What I have said thus far applies to very simple systems of differential equations as well as complex ones. The solutions of the equations may be unstable but they do not seem to exhibit the kind of behavior we so directly associate with indeterministic or probabilistic behavior. It might be argued that the simplest systems of linear differential equations that are unstable do not represent something comparable to indeterminism. Yet it is true that for such unstable linear systems the accuracy of predictions will be poor, given, as is always the case in real situations, any errors in the measurement of initial conditions. In other words, unstable deterministic linear systems capture an important aspect of indeterminism, namely, our inability to predict future behavior on the basis of knowledge of present behavior. There is another aspect also of such linear systems that needs to be noted. In most applications, the linearity of the real system that is being modeled by the linear differential equations is only approximate.

Almost always, deviations from linearity in the real system $\mathcal{J}$—the fact that the linear differential equations are only approximations—will make our ability to predict actual phenomena even more limited.

3. CHAOS AND SYMBOLIC DYNAMICS

We now get down to essentials. Those special unstable solutions of differential equations that exhibit chaotic behavior provide the intended alternative to indeterminism. It would have been more accurate in certain ways to entitle this lecture 'Indeterminism or Chaos, Does it Matter?,' but the meaning of chaos is too special, and so it is the central concept of instability that should be kept to the fore.

So, what do we mean by **chaos**? A brief but not quite technically correct definition is the following. A solution of a deterministic system of differential equations is chaotic if and only if it exhibits some aspect of randomness—or, as an alternative, sufficient complexity. To some, this definition would seem to embody a contradiction, and therefore no solutions would satisfy it. On the left-hand side we refer to a deterministic system of equations and on the right-hand side to the random character of its solution. How can a deterministic system have a random solution? This is what chaos is all about, and the discovery of the new phenomena of chaos is certainly a watershed change in the history of determinism.

Before turning to the recent discussions of chaos, it will be useful to go back over the earlier history of developing the theory of random processes within classical mechanics. The origin of the approach, usually called the method of arbitrary functions for a reason to be explained in a moment, originates with Poincaré, but has been developed in detail by a number of mathematicians in the first half of this century. Already a rather short qualitative sketch of the ideas in very accessible form is given by Poincaré in *Science and Hypothesis* (1913). (The history of developments since Poincaré has been chronicled in some detail by von Plato (1983).) Here I shall just give a sketch of the analysis of coin flipping, one of the most natural cases to consider. To a large extent I shall follow the recent treatment due to Keller (1986), but as somewhat modified in Suppes (1987). Without going into details, we shall assume a circular coin that is symmetric in all the ways you would imagine; second, dissipating forces of friction are entirely neglected; third, it is assumed that the coin does not bounce but on its initial point of impact flattens out to a horizontal position. In other words, from the initial point of impact the face up does not change. With this idealized model, the physical analysis is simple. Newton's ordinary law of gravity governs the vertical motion of the particle—we assume there is no horizontal motion. Second, we assume that the rotational motion is that of constant angular velocity so there is no angular acceleration to the rotation. Now with this situation, if we knew the exact initial conditions, we could predict exactly how the coin would land, with either heads or tails face up. In fact, the classical analysis of this case assumes rightly enough that we do not know the exact value of the initial conditions. The method of arbitrary functions refers to the fact that we assume an *arbitrary* probability distribution of initial vertical velocity and initial rotational velocity. Then as the initial velocity tends to infinity, whatever the arbitrary distribution we begin with, the probability of a head will be one-half. In other words, the symmetry in the mechanical behavior of the system dominates completely as we approach the asymptotic solution. Of course, in real coin-flipping situations we are not imparting an arbitrarily large vertical velocity to the coin, but the variation in the way that we flip will lead to a very good approximation to one-half. The point is that in this typical analysis, the randomness enters only through the absence of knowledge of initial conditions. It is an important example of randomness in mechanical systems, one that has only recently begun to be recognized again as an important example, but it is not the kind of example on which I want to concentrate here.

To show that the conventional philosophical dichotomy between determinism and randomness is mistaken, I consider two important and much discussed examples.

The first is a special case of the three-body problem, certainly the most extensively studied problem in the history of mechanics. Our special case is this. There are two particles of equal mass moving according to Newton's inverse-square law of gravitation in an elliptic orbit relative to their common center of mass which is at rest. The third particle has a nearly negligible mass, so it does not affect the motion of the other two particles, but they affect its motion. This third particle is moving along a line perpendicular to the plane of motion of the first two particles and intersecting the plane at the center of their mass. From symmetry considerations, we can see that the third particle will not move off the line. The restricted problem is to describe the motion of the third particle. The analysis of this easily described situation is quite complicated and technical, but some of the results are simple to state in informal terms and directly relevant to my focus on determinism and randomness. (A more detailed discussion is given in the preceding article on propensity in this volume.)

What can be shown is that any random sequence of heads and tails corresponds to a solution of the deterministic differential equation governing the motion of the third particle. In other words, for each random sequence there exists a set of initial conditions that determines the corresponding solution. Notice that in essential ways the motion of the particle is completely unpredictable even though deterministic. This is a consequence at once of the associated sequence being random. It is important to notice the difference from the earlier coin flipping case, for no distribution over initial conditions and thus no uncertainty about them is present in this three-body problem. No single trajectory in the coin-flipping case exhibits in itself such random behavior.

This example demonstrates the startling fact that the same phenomena can be both deterministic and random. The underlying explanation is the extraordinary instability of the deterministic phenomena.

Before remarking further on the significance of this result, I turn to the second example which is an abstract discrete model of period doubling. Because the mathematics is more manageable it is a simple example of a type much studied now in the theory of chaos. The example also illustrates how a really simple case can still go a long way toward illustrating the basic ideas. Let f be the doubling function mapping the unit interval into itself.

$$(1) \qquad x_{n+1} = f(x_n) = 2x_n (\text{mod } 1),$$

where mod 1 means taking away the integer part so that x_{n+1} lies in the unit interval. So if $x_1 = 2/3, x_2 = 1/3, x_3 = 2/3, x_4 = 1/3$ and so on periodically. The explicit solution of equation (1) is immediate:

$$(2) \qquad x_{n+1} = 2^n x_1 (\text{mod } 1).$$

With random sequences in mind, let us represent x_1 in binary decimal notation, i.e., as a sequence of 1's and 0's. Equation (1) now can be expressed as the rule: for each iteration from n to $n+1$ move the decimal point one position to the right, and drop whatever is to the left of the decimal point:

$$.1011... \rightarrow .011... \ .$$

We think of each x_n as a point in the discrete trajectory of this apparently simple system. The remarks just made show immediately that the distance between successive discrete points of the trajectory cannot be predicted in general without complete knowledge of x_1. If x_1 is a random number, i.e., a number between 0 and 1 whose binary decimal expansion is a random sequence, then such prediction will be out of the question unless x_1 is known. Moreover, any error in knowing x_1 spreads exponentially—the doubling system defined by equation (1) is highly unstable. Finally, it is a well-known result that almost all numbers are random numbers in the sense defined.

Although the exact technical details are rather complicated for almost all chaotic systems, the first example of a restricted three-body problem was meant to illustrate orbital complexity and the second complexity of initial conditions. In any case, randomness can be an essential part of the behavior of what seem to be quite simple deterministic systems.

4. THE TROUBLESOME CASE OF QUANTUM MECHANICS

From what I have just said, the elements of a rejoinder to my earlier skepticism about determinism are apparent. The phenomena cited as examples of indeterminism are in fact just examples of highly complex, unstable deterministic systems whose future behavior cannot be predicted.

The strongest argument against such a view comes from quantum mechanics. Beginning in the 1930s there has been a series of proofs that deterministic theories are in principle inconsistent with quantum mechanics. The first proof of the impossibility of deterministic hidden variables was by von Neumann. The latest arguments have centered on the inequalities first formulated in 1964 by John Bell. Moreover, the associated experiments that have been performed have almost uniformly favored quantum mechanics over any deterministic theory satisfying the Bell inequalities. To those who accept the standard formulation of quantum mechanics, the various proofs about the nonexistence of hidden variables answer decisively the question in the title of this lecture. Indeterminism or instability, does it matter? For these folk the answer is affirmative. The negative results show chaotic unstable deterministic mechanical systems cannot be constructed to be consistent with standard quantum mechanics. The conclusion of this line of argument is that standard quantum mechanics is the most outstanding example of an intrinsically indeterministic theory.

There is, however, a still live option for those of us who are not entirely happy with the orthodox theory of quantum mechanics and its many peculiar features. The option left open is to account for quantum phenomena in terms of something like the theory of Brownian motion, which is, of course, part of classical mechanics broadly construed. Nelson (1967, 1985) has provided thus far the best defense of this approach. He has, for example, derived the Schroedinger equation, the most important equation of nonrelativistic quantum mechanics, from the assumptions of Newtonian mechanics. However, his recent analysis (1985) ends up with Bell's theorem and the relevant experiments as a serious problem. The most feasible way out seems to be to develop a non-Markovian stochastic mechanics, which in itself represents a departure from classical nonlocality. The central problems of current physics are not much concerned with this alternative, but mathematicians and philosophers will continue to puzzle over the foundations of this century's most successful scientific theory. As long as the stochastic view in the sense of Brownian motion remains a viable option, the question posed in the title can be answered by a skeptical "Perhaps not." Consistent with this view, Laplace's concept of probability and thus of indeterminism also remains a viable option—probability is the expression of ignorance of deterministic causes.

5. RANDOMNESS AS A LIMITING CASE OF UNSTABLE DETERMINISM

The existence of deep-seated randomness inside deterministic systems can be attributed to their great instability, and this suggests the road of rapprochement between determinism and randomness. A striking feature of randomness and instability is complexity. Moreover, recent definitions of randomness are in terms of complexity. The complexity of a sequence of finite symbols is measured by the length of a minimal computer program that will generate the sequence. (For asymptotic purposes, the particular computer or computer language does not matter.) A simple alternating sequence of 1's and 0's can be generated by a very short program. More intricate sequences require longer programs and are therefore more complex. Where this argument is going should be apparent. Random sequences are of maximal complexity. In fact, the programs required to generate them would have to be infinitely long. So what are random sequences? They are the limiting case of increasingly complex deterministic sequences. Randomness is just a feature of the most complex deterministic systems. And what of particular importance follows from this? The separation of determinism and predictability. The most complex deterministic systems are completely unpredictable in their behavior. Laplace's "higher intelligence" must be transfinite. He must be able to do arbitrarily complex computations arbitrarily fast. To give a modern ring to Laplace's basic idea, I propose this. *Randomness is the expression of maximally complex deterministic causes.*

6. DOES IT MATTER?

Setting aside, for the moment, the problem of hidden variables in quantum mechanics, we may argue that the philosophically most interesting conclusion to be drawn from the analysis outlined in this paper is that we cannot distinguish between determinism and indeterminism.

The true-blue determinist can hold, without fear of contradiction, that all processes are determined. Confronted with the myriad examples of natural phenomena that cannot be predicted and that seem hopeless to try to predict, he can reply with serenity that even these processes are deterministic, but they are also unstable. The determinist can agree amiably enough that there are processes yet to be analyzed and that his belief that they too will turn out to be deterministic is only based on past experience. This last remark is meant to ring a Bayesian bell. Pure Bayesians are natural true-blue determinists. After all, de Finetti begins his two-volume treatise on probability by printing in capital letters: PROBABILITY DOES NOT EXIST, a thesis Laplace would have heartily endorsed.

The indeterminist, for his part, can just as firmly hold on to his beliefs, directly supported as they are by the phenomenological data in so many areas of experience.

Moreover, with the possible exception of quantum mechanics, there seems to be no current possibility of giving a knock-down argument for either determinism or indeterminism. Under either theoretical view of the world, most natural phenomena cannot be analyzed in detail, and even less can be predicted. How drastic and serious these limitations are is not sufficiently appreciated. I gave a number of obvious examples in the first section, but even in that presumed citadel of mathematically developed science, classical mechanics, it is beyond our current capabilities to analyze a general system of one particle having a potential with just two degrees of freedom.[2]

Whichever philosophical view of the world is adopted, the impact on theoretical or experimental science will be slight. Probability has a fundamental role no matter what, and statistical practice is complacently consistent with either determinism or indeterminism. (The assumption of determinism plays no systematic role in Bayesian statistics, for example.)

There remains the question of whether proofs of no hidden variables in quantum mechanics make a decisive argument against classical determinism. I have mentioned already some reasons for not accepting these results as the last word. I want to conclude with a more general argument. The essential point is the exceedingly thin probabilistic character of quantum mechanics. Roughly speaking, no correlations or other interactive measures can be computed in quantum mechanics. Perhaps most important, if we are examining the trajectory of a particle, no autocorrelations can be computed, i.e., correlations of position at different times, but such a statistic is a most natural measure of probabilistic fluctuation in the temporal behavior of a particle. The probabilistic gruel dished

[2] A system of one particle with two degrees of freedom is a system defined by the differential equations $\ddot{x} = f(x)$, where x is a vector in the plane and f is a vector field on the plane. The system has a potential if there is a function U from the plane to the real numbers such that $f = -\partial U/\partial x$.

out by the wave function of a quantum-mechanical system is too thin to nourish any really hearty indeterminist. Paradoxically enough, the reconstruction carried out so far of quantum phenomena within classical mechanics is probabilistically much richer. It would be ironical indeed if the deepest probabilistic analyses of natural phenomena turn out to be within a deterministic rather than indeterministic framework.

PART IV

PHYSICS

19

DESCARTES AND THE PROBLEM OF ACTION AT A DISTANCE

1. INTRODUCTION

My aim in this Note is to examine Descartes' position on the problem of action at a distance. Since the time of ancient Greece philosophers and physicists have puzzled over the phenomena which seem to show that one body can act upon another at a distance. Many have proposed to solve the problem by introducing sufficient kinds and quantities of unobservable matter to reduce every appearance of action at a distance to a series of contiguous actions; but they have been unable to silence the skepticism of those who could find no independent evidence for the existence of this new matter. On the other hand, those who have erected action at a distance itself as an ultimate principle have been unable to convince their fellow-investigators that it cannot be eventually explained away by more satisfactory modes of contact action. The result has been an interminable controversy still unsettled in our own time.[1]

Descartes' handling of the problem is particularly interesting because of the enormous influence of his general ideas on the history of physics, particularly in the seventeenth century (Mouy, 1934; Bouillier, 1968; Whewell, 1857, II, 102-108, 151-57).

*Reprinted from *Journal of the History of Ideas*, **15** (1954), 146-152.

[1] For example, in the area of electromagnetic phenomena, where contact-action theories have long held sway, action at a distance has been revived by J. A. Wheeler and R. P. Feynman, (1945).

His most systematic statement on physics is his *Principia Philosophiae*[2], first published in 1644. Although the main outlines of this work are too familiar to require re-statement here, it is appropriate briefly to examine Descartes' attempt to reduce the whole of physics to kinematics. He thought that this reduction was effected by his reduction of the concept of body to that of geometrical solid and by his purely relational definition of motion. The clear and distinct notions of size, figure and motion are adequate for explaining everything concerning physical bodies, provided that the principles of mathematics and geometry are accepted.[3]

The quantity of motion is then defined as the product of the size (magnitudo) of a body and its velocity (as a scalar only).[4] Within this framework, a kinematical concept of force is defined. Force is simply the quantity of motion. This definition is not given explicitly and formally, but is easily deduced from repeated uses of the word in certain contexts.[5]

"After having examined the nature of motion, it is necessary that we consider the cause of it."(*Principia*, II, Art. 36). At this point the kinematical program is abandoned and dynamics is introduced. It is to be noted in this connection that the concept of cause is not analyzed. Descartes explicitly requires that the particular causes of motion be clear and distinct, but he apparently tacitly assumes that the concept of cause itself has these two characteristics. This is also true of the more particular dynamical concepts of force and action. Particular forces must be clear and distinct, but the dynamical concept of force as the cause of motion is not formally considered. This second use of the word "force" may, however, like the kinematical use, be easily deduced from the contexts in which it occurs. (*Principia*, II, Art. 25, Art. 26, Art. 37, Art. 43, Art. 57-61, Art. 63.) In this sense, "force" is synonymous with "physical cause of motion." (Descartes' restriction of the physical causes of non-uniform motion to impact forces is discussed below.) Like " force," "action" is also used informally and with its ordinary, commonsense physical meaning.

[2] Descartes, R. Oeuvres, Adam and Tannery Ed. (Paris, 1897), VIII.

[3] *Principia*, IV, Art. 203; see also II, Art. 64, IV, Art. 199. For comment on this point, see K. Lasswitz, *Geschichte der Atomistik* (Hamburg, 1890), II, p. 97.

[4] It is sometimes held that Descartes defined what is ordinarily considered as momentum, but this is a definite error, for he had no proper notion of mass. That by "magnitudo" Descartes simply meant size or volume and not mass is supported by passages in the following articles of *Principia*, II, Art. 36, Art. 40, Art. 43, Art. 47-52, IV, 199, 203. The quantity he does define is nearly useless. It does have two virtues: it is consistent with Descartes' kinematical viewpoint, and it pointed the way toward a correct definition of momentum.

[5] *Principia.*, II, Art. 47-52. Cf. Spinoza, *The Principles of Descartes' Philosophy*, (1943), p. 88, "It should be noted here, that by force (vis) we understand the quantity of motion."

(*Principia*, II, Art. 25, Art. 26, Art. 29, Art. 49, Art. 53, Art. 56, IV, Art. 15-28.)

If the actual development of Descartes' theory had been limited to kinematics, the problem of action at a distance would have assumed a peculiar meaning. However, since he does use the dynamical concepts of force and action with their ordinary physical meaning, the problem assumes its traditional significance. His answer to the problem is well-known, but in order to see its full import we want to make clear the fundamental epistemological distinction between Parts II and III of the *Principia.*

2. THE A PRIORI AND THE HYPOTHETICAL FOR DESCARTES

By contrasting the principles of Parts II and III, and their epistemological status, we are quickly led to a decision as to whether Descartes met the problem of action at a distance on a priori or a posteriori grounds. The general principles of material things, which comprise Part II, are all a priori, and the list of these principles is surprisingly large: they range from a denial of the existence of atoms to a statement of the law of inertia. For the validation of these many principles no appeal to experience is required. In fact, we must be careful not to be deceived by our senses, for our senses do not teach us the true nature of things but only that things are useful or hurtful. The procedure is to "rely upon the understanding alone, by reflecting carefully on the ideas implanted therein by nature." (*Principia*, II, Art. 3). The results of Part II rest upon clear and distinct ideas and are therefore certain. No evidence of our senses could be used to disprove them; no experiments could be performed to refute them. (*Principia*, III, Art. 4).

On the other hand, in Part III, when Descartes turns to consideration of the visible world, he admits that pure deduction from the certain, a priori principles developed in Part II is not sufficient to account for the actual phenomena of experience, in this case especially the motion of the heavens. The principles of Part II are necessary but not sufficient to account for these phenomena. (*Principia*, III, Art. 4). The explicit consideration of phenomena or experiments, a switch from pure rationalism to at least a partial empiricism, is thus necessary to account for the details of the natural world. We are able to know by force of pure reason neither the size of the parts into which matter has been divided, the velocity of these parts, nor their paths. These things could have been ordained by God in an infinite number of different ways.[6] In order to account for the

[6] In other words, the system of a priori principles of Part II is non-categorical, in the

world as it now appears to us, we are thus free to make hypotheses about how God originally ordered the various parts. We merely require of any such hypothesis that its consequences must be in accord with experience. (*Principia*, III, Art. 46). We may in fact know that our hypotheses are false, because of some revealed truth of religion for example, but that does not prevent their being useful and functioning as if true to permit the arrangement of natural causes to produce desired effects. (*Principia*, III, Art. 44, Art. 47). As the particular fundamental hypothesis to account for the phenomena of the visible world Descartes introduces his famous vortex theory.[7] The fundamental assumptions of this theory are: 1) there is order rather than chaos at the beginning; 2) the parts of matter are all equal and moderate in size and velocity; 3) each part has two motions, rotation around its own center and movement with other parts around some fixed center. The motion around the fixed centers provides the only macroscopic inequality in an otherwise isotropic universe (*Principia* III, Art. 46, Art. 47).

We may now ask: if Descartes' general theory of matter and motion is irrefutable by experience, and if, on the other hand, the vortex theory and its consequences are *refutable*, at what point is action at a distance rejected? The answer must be that the principle of contact action is part of the a priori knowledge that is independent of the evidence of the senses. The phrase "actio in distans" does not, I believe, occur anywhere in the *Principia*.[8] The result is that the explicit rejection of action at a

sense that these principles may hold in two different worlds which are not isomorphic.

[7] Stock (1931) emphasizes the role of hypothesis in Descartes' physical thought. Kahn (1918) has a similar thesis. Kahn argues that Descartes began his work as a naturalist, demanding that we go to experience to get answers and that we examine empirical evidence rather than the dicta of authority as a basis for reaching conclusions. However, due to the religious conservatism of his time, Kahn argues, Descartes was forced by external pressure to introduce deductive, a priori methods and to integrate God into physics. Most commentators, such as Lasswitz, would question this thesis, I believe. Descartes' deductive procedures are too thoroughly a part of his method to have been completely put upon him by Church pressure (Lasswitz, 1890, II, pp. 55-57). On the other hand, it is no doubt true that the strong rationalistic tendency of Descartes' thought has been traditionally overly emphasized.

[8] References to action at a distance can be found scattered throughout his other writings. Descartes was particularly concerned to give an explanation of gravity which would avoid any reference to occult forces, i.e., forces which either assume an inherent attraction between distant bodies or act in a manner similar to the action of the soul. A few examples are the following. In a letter to Mersenne, 13 July 1638, he examines three possible explanations of gravity and explicitly rejects attraction as admissible. (*Oeuvres*, Adam and Tannery Ed., II, pp. 223-224). In a letter to Princess Elizabeth, (III, p. 667), he asserts that gravity, heat, etc., are not substances distinct from body and he does not see how attraction would work as a mechanism (III, p. 667). In another letter to an unknown correspondent he asserts the cause of gravity is neither a real quality nor some attraction of the earth, (I, p. 324). In yet another passage he

distance must be constructed as an inference by the reader. However, this inference is not a difficult one, and I imagine no serious reader has ever misunderstood Descartes' position on this matter. In stating the three "laws of nature," which are a priori, Descartes commits himself entirely to impact forces and thus to a clear, although tacit, rejection of action at a distance. The first law asserts that every body continues in the same state as long as possible and that it is changed *only* by colliding with other bodies (*Principia*, II, Art. 37). Any kind of effective action at a distance is rejected by the use of "only." This law is known a priori because God is immutable and always acts in the same way. The second law of nature is that all bodies which are moved tend to continue their movements in a straight line. This is only violated when they *meet* other bodies (*Principia*, II, Art. 39). This law, like the first, is deduced from the immutability of God and the fact that He conserves the motion of matter in the simplest possible way. It also entails the a priori rejection of all attractive or repulsive forces acting at a distance and causing a body to deviate from a state of uniform motion, for such forces are not a case of collision as required by the law. The third law asserts that if a body meets another which has a greater quantity of motion, it loses none of its motion, but if it encounters one having less quantity of motion, it loses as much motion as it transmits to the latter (*Principia*, II, Art. 40). Descartes goes on to say that all the particular corporeal causes changing the state of a body are comprised in this rule, and thus again any action at a distance is ruled out, for such action would not be a case of bodies colliding and could not, therefore, be subsumed under this rule.

Moreover, after the proof of the third law Descartes declares that the force of each body simply consists in the inertia of each body to remain in the same state of motion (*Principia*, II, Art. 43). Through this force of inertia a body may act on another by impact, and may in turn resist the impact of another body. Descartes emphasizes that the force of a body consists only in this inertial property; it has no active attractive or repulsive powers of any kind. Thus we see that to his a priori kinematics, Descartes added but one kind of dynamical force, that of impact.[9] Every

says that to endow particles with the power of acting at a distance would make them "vraiment divines," (IV, p. 396).

[9] It should be noted that Descartes has traditionally been severely criticized for his inadequate account of how a body possessing only the property of extension could have resistance to impact. In the *Principia* this is an untouched mystery. In correspondence with Henry More, Descartes states the following view: "It cannot be understood that one part of an extended thing penetrate another equal to it without the middle part of that extension being, by that fact, destroyed, or annihilated; but what is destroyed does not penetrate the other; and, so, in my judgment, it is demonstrated that impenetrability belongs to the essence of extension, and not of any other thing." *Oeuvres*,

change in the state of motion of a body is to be accounted for by the impact of other bodies upon it, or, what amounts to the same thing, by its impact on them. The contact action of impact is the only mode of action between material things which is clear and distinct. Descartes' a priori mechanics is nearly faithful to his program, for this single dynamical cause of motion is conceived in terms of the figure, size and motion of bodies. Descartes emphasizes this by asserting that the quantity of inertial force of a body is a function of the size of the body, the surface which separates it from other bodies, and the speed of its motion.

Other a priori arguments of Descartes which logically entail the rejection of action at a distance can easily be given, but these centering around the three laws of motion are sufficient to show how completely he accepted the principle of contact action on a priori grounds. The consequence is that the explanation of every phenomenon must, without exception, be given in terms of a mechanism of contact action. As E. T. Whittaker has remarked (Whittaker, 1910, p. 3), this places a heavy burden upon Descartes' system. The explanation of gravitation, light, heat, fire, magnetism and the motion of the planets must in each case involve a mechanism of impact or pressure. Every hypothesis which is made to account for any of these phenomena must use contact action, and the often disastrous results of adopting such a priori principles of natural knowledge are nowhere better illustrated than in Descartes' detailed explanations in Parts III and IV of the physical phenomena mentioned above.

3. CRITICAL REMARKS

This analysis of Descartes' position on action at a distance leads to at least three major criticisms of his physical theory as expounded in the *Principia*.

1. A mechanics based on a priori principles seems doomed to failure, for principles which are above experience, unalterable and irrefutable, can never be abandoned for principles more in conformity with empirical observations. Historically, there is an interesting parallel between Descartes and Kant. Kant's solution of the problem of action at a distance was different from Descartes', for he made both a principle of contact action and a principle of action at a distance a priori synthetic.[10] Methodologically, however, the two philosophers stand together, for they both offered a solution of this fundamental physical puzzle on a priori grounds.

Adam and Tannery ed., V, p. 378.

[10] *Metaphysische Anfangsgrunde der Naturwissenschaft* (Riga, 1786), Zweites Hauptstück.

Kant's analysis is, of course, more sophisticated than Descartes', but it suffered the same fate: incompatibility with the later development of physics.

It has been argued that if it had not been for the historical accident of Newton and the relative weakness of contemporary Cartesian physicists, the Cartesian physics might have been corrected and further developed (Mouy, pp. 321-322). This is a defensible speculation if only the *results* of the Cartesian physics are considered. If, however, the *methods* by which these results were validated are also considered, it does not seem defensible. It is true that the hypothetical, refutable vortex theory could have been changed without violating basic Cartesian tenets, and thereby some of the detailed explanations of particular phenomena could have been considerably improved. However, the same kind of tampering with the fundamental mechanical principles set forth in Part II of the *Principia* could not have been tolerated. From Descartes' standpoint, to deny seriously the truth of any principle stated in Part II would have been as absurd as to deny the truth of a theorem of Euclid, for every one of these principles belongs to the domain of mathematics and geometry (*Principia*, II, Art. 64). Since the principles of Part II are noncategorical, that is, do not uniquely determine the complete structure of the physical world, there exists the logical possibility of supplementing them by new hypotheses replacing the vortex theory. Nevertheless, it is unlikely that a set of hypotheses could have been found which would have been both empirically adequate and logically consistent with the a priori principles.

2. Although in a general sense the hypothetical methods of Parts III and IV of the *Principia* are acceptable, the actual analyses of particular physical phenomena are almost completely unsatisfactory. What is the main reason why the developments in these parts of the treatise now seem so ridiculous, particularly when compared with the physical treatises of Galileo and Newton? The central weakness, it seems to me, is the wholesale postulation of unobservable particles which are assigned complicated, yet purely qualitative, imprecisely defined structures. The various microscopic particles introduced by Descartes are all slavishly modeled after the macroscopic bodies observed and encountered in ordinary experience. This obviously inadequate method of analogy is the main technique employed in passing from the general vortex theory to particular phenomena.

A second, closely related weakness of Parts III and IV is that the logical consequences of the many subsidiary hypotheses are not pursued in any detail. This failing led to the early downfall of Descartes' system in the most important and precise branch of seventeenth-century science, namely, the mechanics of the solar system. The vortical explanation of the motion of the planets was demolished by Newton.[11]

3. Descartes' use of an ideal fluid on the one hand, and of shaped particles on the other, to constitute the plenum is one of his most serious and fundamental confusions. There seems to be a clear reason why this or some other comparable inconsistency was inevitable in his system. The hydrodynamics of an ideal, non-viscous fluid seems to be the natural physics of contact action, and this is the physics dominating Part II. A particle of such a fluid is essentially a point without figure or size; it cannot have definite shape, size or rigidity. Consonant with this physics of fluids, Descartes denied on a priori grounds the existence of atoms (*Principia*, II, Art. 34,35) and the existence of attractive and repulsive forces acting at a distance. In thus limiting his physics so severely, he effectively eliminated any device for explaining the specific variety of bodies encountered in experience. The result was that when he turned from the general theory to the explanation of particular phenomena, he was forced to introduce surreptitiously either shaped particles, i.e., atoms, or dynamical forces of attraction and repulsion. The course that he does adopt is the one least inconsistent with his position on action at a distance. The shaped particles are made to explain observed phenomena by their actual motions, which can only be changed by impact with other particles. Dynamical forces of attraction and repulsion, existing independently of the actual motions of the particles, are rejected as thoroughly in Parts III and IV as in Part II.[12]

This general mechanical ideal of *reducing* the particular causes of all changes in nature to simple cases of contiguous forces of impact has exercised an enormous hold on the development of physics even until recent years. Yet from a systematic philosophical standpoint, Cartesian physical theory is an example of reductionism at its worst. This reductionism is probably the source of the paradoxical historical position of Descartes' physics: his simplifying general ideas had great influence, yet his positive technical contributions were slight. A physics based on a very few clear ideas is perennially appealing, but it can be empirically sound and technically interesting only if provided with a powerful mathematical frame-

[11] Newton simply pursued the logical consequences of Descartes' hypotheses far enough to show that they were inconsistent with Kepler's second and third laws, and could not account for the observed motions of comet or planetary satellites. *Philosophiae Naturalis Principia Mathematica*, Cajori translation pp. 395, 396, 543.

[12] It is interesting to note that Boscovich and Kant took the opposite course: they denied the existence of atoms and affirmed the existence of dynamical forces. In the working out of details, neither was as inconsistent as Descartes.

work, which is precisely what Descartes did not provide for his theory. Indeed, from the standpoint of physics, we may say of Descartes what Locke said of himself: that he served "as an under-labourer in clearing the ground a little, and removing some of the rubbish that lies in the way to knowledge."

20

SOME OPEN PROBLEMS IN THE PHILOSOPHY OF SPACE AND TIME

Philosophical analysis and speculation about the concepts of space and time are as old as philosophy itself. Concurrent with the astounding technical development of greek mathematical and observational astronomy, a polished and carefully articulated theory of space and time was set forth early in the Hellenistic period by Aristotle, especially in the *Physics*. Aristotle's *Physics*, Euclid's *Elements* and Ptolemy's *Almagest* form a triad that elaborate the philosophical, mathematical and physical foundations of space and time in ancient philosophy. Although Ptolemy was Aristotelian in his philosophical attitudes, a clear divergence between Aristotle on the one hand and Euclid and Ptolemy on the other is obvious. These two quite distinct traditions, one Aristotelian and philosophical, and the other mathematical and Euclidean or Ptolemaic, continued in the succeeding millennium and a half leading up to the outburst of modern science in the seventeenth century.

The separate life of the two traditions did not stop even there. Descartes' *Principles of Philosophy*, for example, is much closer in spirit to Aristotle's *Physics* than to Euclid's *Elements* or Ptolemy's *Almagest*. In spite of the fact that in other domains Descartes was a creator of new

*Reprinted From, *Synthese*, **24** (1972), 298–316.

mathematical concepts, in his *Principles* there is no genuine mathematical organization or development of ideas. From a philosophical standpoint, it is evident that the closeness of argument characteristic of Aristotle's *Physics* is not matched, and there is a general degradation of intellectual standard.

The continuation of the Euclidean-Ptolemaic tradition is quite otherwise. Newton's *Principles*, first published in 1687, is very much in the spirit of Ptolemy's *Almagest* and satisfies a standard of intellectual rigor and clarity that would have been acceptable in Alexandria in Ptolemy's time. In Newton's *Principles* there is no sharp separation of mathematics and physics, or of mathematics and astronomy. To a large extent this fusion of mathematical and astronomical investigations was continued a hundred years later in Laplace's *Celestial Mechanics*.

The separation of mathematical investigations on the one hand, and physical or astronomical investigations on the other, did not really occur until the nineteenth century. By the latter half of that century there were very few examples of individuals making original contributions both to mathematics and to physics. Separation of the intellectual traditions was nearly complete by the beginning of this century. I perhaps need to be more explicit in defining this separation. Certainly physicists continued to use mathematics and to use it with great power and sophistication, but original contributions to the foundations of mathematics and original contributions to the conceptual foundations of physics were not made by the same people. Of course a small number of individuals like von Neumann and Hermann Weyl made significant contributions to both domains, but still the generalization is, I think, a sound one, and midway through the last half of the twentieth century it is more valid than earlier.

This scientific separation has given rise to a separation within philosophy, so that to a large extent the philosophical foundations of mathematics, including the foundations of geometry, are now an almost totally separate subject from the philosophy of space and time. In this article I would like to describe some open problems in the philosophy of space and time that require the methods characteristic of mathematical traditions in the foundations of geometry for their solution, and thereby to encourage within philosophy a fusion of the two traditions.

I have organized the analysis of problems under two main headings. In the first section I am concerned with the geometry of space and deliberately deal with classical questions that do not take into account the theory of relativity. In the second section I turn to physical space and space-time, including such classical problems as the formulation of an adequate theory of bodies.

1. GEOMETRY OF SPACE

Because of extensive development, the foundations of geometry are a natural proving ground for more general philosophical concepts in the philosophy of science, but surprisingly little has been done to use this proving ground. I restrict myself to two classes of problems. The first class deals with the attempt to give operationalism a sharp foundation in the case of the measurement of spatial relations or, more generally, in terms of the geometry of space. The second class deals with combining measurement and error to yield some systematic theory of approximation. The idea of such approximations is familiar both in physics and in psychology. What is not familiar in either discipline is the development of geometrical foundations of such approximations.

Operational foundations. Compared with the notion of constructivity in the foundations of mathematics, there have been few attempts to give a sharp formulation of the concept of an operational definition, or of operationalism, in the philosophy of science. The foundations of geometry provide an excellent place to give such a formulation. In the first place, perhaps the oldest issues concerning constructivity in mathematics are to be found in the foundations of geometry. Certainly the three classical problems of Greek elementary geometry—squaring a circle, duplicating a cube and trisecting an angle—are examples of constructive problems. The beautiful thing about these problems is that we can approach the foundations of geometry in a qualitative way, but with the objective of providing a precise solution to the problems. Such a solution, of course, is negative for elementary operations. Problems of a comparable sort have not been formulated in the foundations of physics, but I do not explore that aspect of the problem in this section. In the present context, I want to concentrate only on the purely geometrical aspects of constructivity.

Reflection on the three classical problems or, at a more mundane level, examination of the propositions in the early books of Euclid's *Elements* suggests that existential statements are always backed up by a highly constructive sequence of operations. From a mathematical standpoint, especially from an algebraic one, the natural idea then is to formulate the qualitative foundations of geometry in terms of operations rather than in terms of relations and existential statements about these relations. From a general philosophical standpoint the problem can be expressed as follows: *Characterize operations that can be performed on spatial points so that from the known properties of the operations, the usual properties of space can be derived.*

A first thought might be that these operational problems are already solved by the standard formulations of axioms for vector spaces, but this is not the case for two reasons. First, the vector spaces themselves are not the same in structure as the ordinary Euclidean spaces, because of the distinguished point of origin. Second, the operations in the vector spaces do not correspond to the operations needed, for example, to solve the problems formulated in Book 1 of Euclid's *Elements*, and it is clear that from a geometrical standpoint operations that are closer to the Euclidean constructions are available with many alternative possibilities open.

Put another way, an operationally satisfactory formulation of the constructive part of Euclidean geometry should be a theory in standard formulation, that is, a theory that is formulated within first-order logic with identity and that is also quantifier-free. The reason for the quantifier-free requirement should be apparent from what has already been said. An existential statement in constructive parts of geometry is misleading, because a specific and definite constructive method of finding the point existentially postulated is known. A conceptual discrepancy exists between the axioms and the methods of construction when a general existential statement rather than a specific sequence of constructive operations is postulated. A reason for insisting on such a viewpoint in geometry is that it is possible to get a thorough understanding of the operational situation in a way that is not at present possible in physics. We can realistically hope to give a theory with standard formalization that fully characterizes constructive Euclidean geometry and that does so in an elementary way. It is not yet clear that we understand how to do this in any thoroughgoing fashion for substantial parts of physics, although this is a topic on which I shall have more to say later.

In an earlier paper (Moler and Suppes, 1968), a constructive formulation of geometry in the sense just defined was given. This formulation depends on two primitive operations: one the operation of finding the point of intersection of two line segments; and the other, the operation of laying off one line segment on another. Both of these operations are discussed in some detail in Hilbert's *Foundations of Geometry*, but our task was to give an explicit axiomatic formulation in terms of just these two operations and in quantifier-free form. The axioms turn out to be complicated, and a simpler and more elegant quantifier-free formulation in terms of other primitive operations is needed. For example, let I be the intersection operation and S the laying-off operation so that $I\ (xyuv)$ is the point of intersection of the line determined by x and y with the line determined by u and v, and $S(xyuv)$ is the point as distant from u in the direction of v as y is from x. Then betweenness is defined by:

$$B(xyz) \quad \textit{iff} \ [\text{if } x \neq z \ \text{ then } S(xyxz) = y = S(zyzx)] \ \&[x = z \rightarrow x = y],$$

collinearity is defined by:

$$L(xyz) \quad \textit{iff} \quad S(xyxz) = y \text{ or } S(zyzx) = y \text{ or } x = z,$$

and noncollinearity of four points is defined by:

$$NL(xyuv) \quad \textit{iff} \quad \text{not}(L\,(xyu) \ \text{ or } \ L\,(yuv) \ \text{ or } \ L(xuv) \ \text{ or } \ L\,(xyv)) \ .$$

Euclid's axiom, the most complicated of the 18 axioms of the system, then has the following formulation:

$$\begin{gathered}\text{if } NL\,(xyuv) \ \& \ B(x, I(x, S(xyuv), y, u), S(xyuv)) \ \& \\ S(y, S(xyuv), \ x, \ u) \ \neq \ u \text{ then } L\,(x, y, I(xyuv)),\end{gathered}$$

which is far from transparent in its content, although we know an axiom of approximately this sort is necessary.

One conjecture is that it is a mistake to take points as the primitive objects. The difficulty with the intersection and laying-off operations formulated in terms of points is that these are quaternary operations, and the properties of quaternary operations as opposed to binary operations are inevitably somewhat complex. In subsequent thinking about the problem, I have looked at axioms based upon the primitive objects' being directed line segments with an operation of addition for such segments. An additional unary operation is that of taking the inverse of a directed line segment. Thus, for example, a line segment plus its inverse yields simply the point of origin of the first line segment. The natural axioms here on addition and the inverse operation are close to those for an additive group as in the case of vector addition, but in the present instance they do not actually satisfy all the axioms. For example, the addition of a directed line segment and its inverse yields not the identity of the group, but the particular point of origin of the first segment. (In fact, we do not even get a Brandt groupoid, because the left-cancellation axiom is not satisfied.) In such a geometry it is also natural to add an operation of a qualitative comparison of length, easily represented by a binary ordering relation. Additional constructive operations, like that of one directed line segment being perpendicular to another, are also easily added. However, I am not satisfied with the full set of axioms I have put together—they are too complicated and again too awkward as in the case of the earlier work.

I am persuaded that with additional effort and insight natural and quantifier-free axioms on simple geometric operations can provide an adequate formulation of constructive Euclidean geometry.

Geometry of approximations and errors. The theory of error in astronomical observations, and more generally in any sort of numerical observations, dates from the work of Simpson, Lagrange and Laplace in the eighteenth century. Their efforts were aimed at problems that arise especially in astronomical observations. Dating from the latter part of the nineteenth century there is also a tradition in psychology concerned with the phenomenon that it is easy to judge, for example, tone A to be just as loud as tone B, tone B to be just as loud as tone C, but tone A to be strictly louder than tone C. This phenomenon of just noticeable differences and the related phenomenon of the nontransitivity of judgments of indifference has received considerable attention, and there is much that is common to the formal theory as applicable to both physics and psychology. However, the extension of the formal theory to spatial concepts, and thereby to geometry, has as yet been inadequately developed, in spite of the considerable conceptual interest in understanding what it is like to have directly a qualitative geometrical theory of error or approximation. About the only qualitative part of the theory that is thoroughly understood is the theory of order. Even then, until the relatively recent discussion by Luce (1956) the problem of formulating the theory of order was not properly considered in explicit fashion. Luce's axioms for semiorders were modified and simplified in Scott and Suppes (1958). The theory is developed for a binary relation in one dimension. In the following definition, I call a binary structure an ordered pair $\mathfrak{A} = \langle A, R\rangle$ such that A is a nonempty set and R is a binary relation on A. The definition of semiorders is then easily given in elementary form.

DEFINITION 1. *A binary structure* $\mathfrak{A} = \langle A, R\rangle$ *is a* semiorder *if and only if the following axioms are satisfied for every x, y, z and w in A:*

1. *Not xRx;*
2. *If xRy and yRz then either xRw or wRz;*
3. *If xRy and zRw then either xRw or zRy.*

The following representation theorem for such semiorders can then be proved.

THEOREM 1. *If* $\mathfrak{A} = \langle A, R\rangle$ *is a finite semiorder, that is, A is a finite set, then there is a real-valued function φ such that for every x and y in A,*

$$\varphi(x) > \varphi(y) + 1 \text{ iff } xRy.$$

The closely related binary relation I of indistinguishability has been thoroughly investigated by Roberts (1970). (We define I as follows in terms of R: *xIy iff not xRy and not yRx.)* The surprising thing Roberts shows is that indistinguishability, unlike semiorders, is not axiomatizable in an elementary fashion by a finite set of open sentences. It is of course clear that a semiorder is not a natural relation in geometry because a direction on the line is assumed, and clearly the binary relation of indistinguishability by itself does not have very much geometrical content, although its topological properties have been developed in a thorough way by Zeeman (1962).

The next natural thing is to ask for order on the line. Classical axioms for betweenness on the line may be stated in terms of a ternary structure that is a nonempty set A and a ternary relation B, interpreted as betweenness. A variant of axioms that may be found in the literature is given in the following definition:

DEFINITION 2. *A ternary structure* $\mathfrak{A} = \langle A, B\rangle$ *is a* one-dimensional betweenness structure *if and only if the following five axioms are satisfied for every* x, y, z *and* w *in* A*:*

1. $yB(xyx)$ *then* $x = y$;
2. *If* $B(xyz)$ *then* $B(zyx)$;
3. *If* $B(xyz)$ *and* $B(ywz)$ *then* $B(xyw)$;
4. *If* $B(xyz)$ *and* $B(yzw)$ *and* $y \neq z$ *then* $B(xyw)$;
5. $B(xyz)$ *or* $B(yzx)$ *or* $B(zxy)$.

On the basis of these axioms, it is straightforward to prove the following theorem.

THEOREM 2. *Let* $\mathfrak{A} = \langle A, B\rangle$ *be a one-dimensional betweenness structure and let* A *be a finite set. Then there is a real-valued function* φ *such that for all* x, y *and* z *in* A

$$[\varphi(x) \leq \varphi(y) \leq \varphi(z) \text{ or } \varphi(z) \leq \varphi(y) \leq \varphi(x)] \textit{ iff } B(xyz).$$

To express the idea of approximation, we can use the notion of ϵ-betweenness, following the developments in Roberts (1973). The intuitive idea is that the relation of betweenness holds to within a small physical or perceptual error. Formally this is caught in the following condition, which replaces the equivalence of the preceding theorem.

$$(1) \qquad |\varphi(x) - \varphi(y)| + |\varphi(y) - \varphi(z)| < |\varphi(x) - \varphi(z)| + \epsilon \textit{ iff } B(xyz).$$

For the formulation of Roberts' axioms we need the additional notion of an indistinguishability relation as discussed above, defined in terms of betweenness: xIy *iff* $B(xyx)$. Of a number of different formulations of indifference graphs given in Roberts (1970), perhaps the simplest one is this. A binary structure $\mathfrak{A} = \langle A, I\rangle$ is an indifference graph *iff* any subgraph, that is, any subset of A, call it A_1, is connected, that is any two points in A_1 are related by some power of the relation I; more precisely, for any x and y in A_1, there is an n such that xI^ny, and any such connected subgraph has at most two extreme points that are not equivalent. (Two points are said to be equivalent if they stand in the relation I to exactly the same points in a graph, and an element e of A is an extreme point if whenever x and y are in A, and both stand in relation I to e, but are not equivalent to e, then x stands in relation I to y, and moreover, there is another element in A that stands in relation I to x and y but not to e.)

The axioms for ϵ-betweenness are then embodied in the following definition.

DEFINITION 3. *A ternary structure* $\mathfrak{A} = \langle A, B\rangle$ *is a* one-dimensional ϵ-betweenness structure *iff the following axioms are satisfied for every* x, y, z, u *and* v *in* A:

1. $\langle A, I\rangle$ *an indifference graph*;
2. *If* $B(xyz)$ *then* $B(zyx)$;
3. *If* $B(xyz)$ *and* $B(ywz)$ *and not* (yIz *and* wIz) *then* $B(xyw)$;
4. *If* $B(xyz)$ *and* $B(yzw)$ *and not* yIz *then* $B(xyw)$;
5. *If* $B(wyz)$ *and* $B(yxz)$ *then* xIy *or* (zIx *and* zIy);
6. *If* xIy *then* $B(xyz)$;
7. $B(xyz)$ *or* $B(xzy)$ *or* $B(yxz)$.

On the basis of this definition Roberts proves the following theorem:

THEOREM 3. *Let* $\mathfrak{A} = \langle A, B\rangle$ *be a one-dimensional* ϵ*-betweenness structure, let* A *be a finite set, and let* $\epsilon > 0$ *be given. Then there is a real-valued function* φ *on* A *satisfying* (1) *above.*

Unfortunately, as is evident from the above axioms, even the theory of ϵ-betweenness is relatively complicated. The axioms in terms of ϵ-betweenness and what we can call ϵ-equidistance, corresponding to the two primitive relations used by Tarski (1959), seem to lead to an extremely complicated set of axioms in order to characterize the 'approximation version' of the Euclidean plane. The problem is open of finding a reasonable set of axioms for the Euclidean plane in terms of ϵ-approximations to standard geometric relations or operations.

2. PHYSICAL SPACE AND SPACE-TIME

In this section I discuss open problems connected with the following topics: the theory of bodies, the operational foundations of special relativity and the conceptual foundations of elementary physics.

Theory of bodies. One program of research investigated from a number of perspectives over many years is that of replacing the classical notion of point or line as primitive concepts in geometry and constructing three-dimensional geometry from the concept of a solid object or body. Fairly extensive efforts in this direction were made, for example, by Whitehead (1919, 1920), who regarded his efforts as a significant application of his method of extensive abstraction.

A brief, but classical, article on this subject is Tarski's 'Approach to the Foundations of the Geometry of Solids,' which takes only Lesniewski's relation of part and the geometrical concept of sphere as primitive. A translation of this work from the twenties may be found in Tarski (1956, pp. 24-29).

The classical tendency has been to impose increasingly strong axioms on bodies in order to obtain ordinary three-dimensional Euclidean space. In Tarski's axiomatization, for example, axioms in terms of the primitive concepts of part and sphere actually play a minor role, for in terms of these concepts he defines the concept of point and the ordinary geometric relations between points.

Of particular philosophical interest is a more restricted theory of bodies. A useful beginning in this direction is provided by Noll (1966). I shall not follow through all of Noll's work, because he extends his axioms to obtain a foundation of mechanics and introduces thereby spatial concepts in an interesting indirect way in terms of representing the force exerted on a body at a given instance by a vector, that is, an element of an ordinary vector space. The initial elementary axioms are close to the ideas of Lesniewski, but almost certainly the theory has been constructed independent of Lesniewski. Noll begins with the relation *part of.* There seem to be good philosophical reasons for substantially changing some of Noll's approach, but the spirit of what I give below draws directly on his work. Although I begin with modified versions of Lesniewski's and Noll's axioms, I add other axioms and concepts that are not at all in the spirit of their developments. What is given here is incomplete and thus perhaps suggestive of some interesting open problems.

Let π be the relation of *part*; in other words, in the intended interpretation $A\pi B$ *iff* A is a part of B. If $B\pi A$ and $C\pi A$, then A is an *envelope* of $\{B,C\}$. Moreover, A is *the least envelope* of $\{B,C\}$ *iff* A is an envelope of $\{B,C\}$ and for any D that is an envelope of $\{B,C\}$, $A\pi D$.

Some additional definitions are useful. Their intuitive content is obvious. A is a *common part* of $\{B,C\}$ *iff* $A\pi B$ and $A\pi C$. Bodies A and B are *separate* *iff* they have no common part. Body A is *a least part* of B *iff* $A\pi B$ and there is no body C such that $C\pi A$ and $C \neq A$. (The clause that $C \neq A$ is required because π is taken to be reflexive and thus every body is a part of itself.) Body A is *the greatest common part* of $\{B, C\}$ *iff* A is a common part of $\{B, C\}$ and for every body D if D is a common part of $\{B, C\}$, $D\pi A$.

We also define partial operations of join and meet. If A is the least envelope of $\{B, C\}$, then $B \cup C = A$, and we say that A is the *join* of B and C. If A is the greatest common part of $\{B, C\}$, then $B \cap C = A$, and we say that A is the *meet* of B and C. The operations are partial, because separated bodies do not have joins and meets.

Finally, let $A_1, \ldots, A_n$ be parts of B, let $A_1 \cup \ldots \cup A_n$ exist, and let $A_1 \cup \ldots \cup A_n = B$, then we say that $\{A_1, \ldots, A_n\}$ is a *finite dissection* of B.

My incomplete set of axioms for bodies is embodied in two definitions. The first six axioms of Definition 4 are a weakened version of Lesniewski's axioms for mereology as formulated in Grzegorczyk (1955), although I use some of the rather natural terminology introduced by Noll (1966). The axioms are weaker than Lesniewski's in that products, sums and differences are not necessarily defined for any two bodies. Stronger conditions are imposed by my axioms for products, sums and differences to exist. These conditions, which seem physically natural, are similar to ones imposed by Noll. For example, for the product or greatest common part of two bodies to exist they must, according to the axioms given here, have a common part. On the other hand, the axioms diverge from Noll's in not postulating the body that is exterior to a given body. The existence of this possibly unlimited exterior seems dubious, and for many intuitive examples, it is not a natural physical object. For instance, the body that is the exterior of the earth or sun is not conceptually well defined in celestial mechanics. The import of the remaining axioms is discussed below.

DEFINITION 4. *A binary structure* $\mathfrak{X} = \langle X, \pi, \rangle$ *is a* structure of bodies *if and only if the following axioms are satisfied for every A, B, C and D in X:*

1. $A\pi A$;
2. *If* $A\pi B$ *and* $B\pi A$ *then* $A = B$;
3. *If* $A\pi B$ *and* $B\pi C$ *then* $A\pi C$;
4. *If A and B have a common part, then they have a greatest common part*;
5. *If A and B have an envelope, then they have a least envelope*;
6. *If A is a part of B and* $A \neq B$, *then there is a body C in X such that B is the least envelope of* $\{A, C\}$;
7. *Every body has a least part*;
8. *Every body has a finite dissection of least parts.*

It should be apparent that it is easy to formulate all but the last of these eight axioms as first-order axioms. For example, Axiom 5 would read:

$$(\exists C)(A\pi C \& B\pi C) \rightarrow (\exists D)(A\pi D \& B\pi D \& (\forall E)(A\pi E \& B\pi E \rightarrow D\pi E)).$$

Axioms 7 and 8 are much stronger and restrictive in character than the first six axioms. They may be regarded as general axioms of abstract atomism. Thus, Axiom 7 might be interpreted as saying that every body contains at least one atom, and Axiom 8 that every body is made up of a finite number of atoms.

There are a number of different ways to extend the axioms of Definition 4, and by heavy-handed methods, we can reach ordinary Euclidean geometry fairly rapidly. We simply have to postulate enough bodies and atoms. We would not of course expect to get the full Euclidean space because of the finite dissection property, but we would want to be able to imbed in three-dimensional space, and to get a representation of this imbedding technique up to the standard group of rigid motions.

There is no doubt that this program can be carried through. The techniques for the one-dimensional case of measurement exploited in many different directions in Krantz *et al.* (1971) provide more than adequate tools, but yet I do not see how to pursue it in a simple and elegant fashion. At the same time I am beginning to see a philosophically interesting aspect of this program if it can be satisfactorily carried through. Properly carried out, it should provide a new way of looking at the nature of space.

For many technical reasons that were clear already in Greek geometry, it is much easier to start with points and to deal with the abstractions that follow not only from consideration of points, but also from consideration of points filling space. It is extremely hard to escape from this way of looking at things. The approaches to geometry that begin with a concept of body or solid, as, for example, those of Whitehead, Tarski, Lesniewski, Grzegorczyk or Noll, end up with a richness of structure that is essentially exactly equivalent to Euclidean three-dimensional space. On the other hand, this is not an idle fact; it must be recognized that we have to come to terms with Euclidean geometry in some form. A theory that does not is obviously too weak to be of serious conceptual interest.

To begin with and to put it baldly, I propose looking at the intuitive concept of space as just a set of possible worlds. Of course, it is a rather special set of possible worlds. It is the set of all possible relative positions of bodies. But insisting on this viewpoint seems to me to clarify a number of problems. Certainly, it strikes down the container theory of space which, in spite of criticisms that go back to Aristotle, continues to be a perennially popular view of space. This viewpoint also gives a deeper analysis of relational theories of space. The difficulty with relational theories is that it is too easy to cast them in terms of actual relations. Rather, we need to think of the set of all possible relations between bodies, and this characterizes space. Where we get in trouble epistemologically is in beginning with points rather than with bodies. It is somewhat like the problem of constructing a sample space in probability theory. We understand the construction of the sample space best when we start with the method of generating the possible sequence of events and use this method of generation to describe the possible experimental outcomes, the set of which constitutes the sample space.

In constructing space as a set of possible relative positions, it is not the concept of point as such that creates difficulties. Rather it is the classical concept of there being so many points. The points ordinarily postulated as existing in space have no more reality under the view advocated here than do the possible sequences in a large number of flips of a coin. The various sequences represent nicely possible experimental outcomes, but in themselves they have no concrete existence. Only one of them will come to represent the actual sequence, and I say the same is true of points. It is not possible here to develop this view thoroughly, but I do think that beginning with the kind of theory of bodies discussed above it is feasible to develop a theory of space from the theory of bodies and to get the concept of space itself out as a construction derived from the set of possible relative positions of bodies.

Moving from positions to trajectories we may obtain a characterization of space-time as the set of all possible trajectories of bodies, and this is probably more fundamental than the separate concept of space.

Special relativity. On several past occasions I have stressed the significance of Robb's axiomatization (1936) of space-time in the sense of special relativity. His axiomatization is important, because of its completeness and the simplicity of its single primitive—the binary relation of *after* holding between space-time events. Robb's important work has been repeatedly ignored by philosophers, but I am happy to say that the long article by Domotor in this volume includes a detailed discussion of Robb's work. The article by Latzer also provides an axiomatic treatment different from, but very close to that of Robb.

As I have remarked in earlier discussions of Robb's axiomatization, the complexity of the axioms stands in marked contrast to the simplicity of his single primitive concept. The point I want to emphasize is the desirability of quantifier-free axioms of the sort discussed above for Euclidean geometry. It is almost paradoxical that no such axiomatizations have yet been given for special relativity. Given the enormous literature on operationalism, its relations to Mach and Einstein, and the extensive discussions of physicists like Bridgman, without knowing the literature one would anticipate that a number of different rigorous treatments of an operational approach to special relativity could be found.

One thing is evident. The kind of primitive operations I discussed earlier for Euclidean geometry do not seem intuitively appropriate for operations in a space-time manifold—I mean the operations of finding the intersection of two line segments and of laying off one line segment on another.

From the results in Suppes (1959b) we should be able to establish a sufficient axiomatic base by considering just segments of inertial paths, because of the invariance of the relativistic measure of such segments. Moreover, as also shown in that article, we use in a natural way parallelogram constructions to get at the relativistic invariance of other segments that are not segments of inertial paths. The explicit proof in Suppes (1959b) of the invariance of such inertial path segments being an adequate basis for deriving the Lorentz transformations requires the use of various elementary geometrical operations, like that of finding a midpoint that could be used in an operational, quantifier-free geometry of special relativity. However, I have been unable to find a transparent way to build up an adequate axiomatic construction from this approach.

The approach begun by Walker several years ago (1948, 1959) may possibly lead to more satisfactory results. Walker takes a richer set of primitives than Robb's, but one's that are related. In addition to events he also has particles, an ordering relation of beforeness on events, and most importantly, a one-one signal-mapping from one particle onto another. With this apparatus, he gives one of the few formal definitions of observables to be found anywhere in the literature of special relativity; namely, an observable is a mapping from the distinguished particle called the observer on to the observer, that is, from that particle on to itself, resulting from a chain of signal-mappings and inverse signal-mappings. My central reservation about Walker's approach is that the signal-mappings are in fact complex functions that do all the work at once that should be done by a painstaking buildup of more elementary operations. At least that is my perspective on the intuitively correct approach. Another remark is that several of his axioms are very powerful; for example, his notion of a particle's being dense makes each particle ordinally equivalent to the continuum of real numbers. All the same, Walker's work, which conceptually derives from the earlier intuitive ideas of Milne, is a clear conceptual alternative to Robb's and marks a distinct advance over the level of rigor and explicitness found in most of the literature.

As the discussion in Latzer's paper in this volume shows, the mathematical problems of finding an adequate qualitative axiomatic basis for the general theory of relativity are complex and formidable. But this is certainly not the case for the special theory of relativity, and it is surprising that so few axiomatic results of a definite nature have as yet been achieved. The absence of such explicit work indicates how poorly we understand in any deep conceptual way the ideas of operationalism that have been current for almost a hundred years. I shall say more about special relativity in the next section on elementary physics.

Elementary physics. Talk about some parts of physics being elementary is fairly frequent, and presumably there is an effort on the part of textbook writers to restrict themselves to that part of physics that is elementary. Actually, the situation is not clear. While a modern secondary-school textbook will probably contain a chapter on quantum mechanics, its discussion is purely qualitative and no actual numerical exercises are worked out.

It is my conviction, reinforced by a number of conversations with Seymour Papert, that the concept of elementary physics can be made an intellectually respectable one, with a precise formulation of what its range of subject matter is. I should make it clear at once that I do not think there is any unique approach to elementary physics; several different ways of formulating the domain are possible. I do think a kind of representation result can be given prominence, and that I want to describe. However, I want to approach that representation theorem somewhat indirectly and begin with a characterization that is natural in the context of the great emphasis on first-order logic in the philosophy of mathematics and science.

One natural approach would be to say that a part of physics is elementary if it can be expressed as a theory with standard formalization in first order logic. Several of the problems discussed earlier in this article have that character, and it is certainly a framework familiar enough in the philosophy of science. Although organizing much of geometry in the first-order framework is easy, it is hard to point to significant examples of physics that have been axiomatized with this restriction. To some extent, this may be due to a lack of sustained effort, and I have the conviction that much real physics can be put within a first-order framework.

Another approach that is closely related but that can get us more quickly into a formulation of several parts of physics, and that is probably at the present time considerably more practical as an actual way of marking off in some systematic fashion elementary parts of physics, is to restrict ourselves to an elementary algebraic approach, in particular, to restrict our field of numbers to an ordered Euclidean field. (An ordered field in the sense of modern algebra is Euclidean if whenever a positive element a is in the field then there is an element b such that $b^2 = a$, i.e., we can take square roots.) We get all the vector space apparatus we need by considering vector spaces over such Euclidean fields, and we then introduce elementary laws of physics by means of special functions which take values either in the field or in a three-dimensional vector space over the field. Simple formulations of the conservation laws of momentum, for example, can easily be made within such a framework.

A second example may be found in the foundations of special relativity as discussed above. It is clear that an elementary geometric foundation can be given for special relativity that has as its representation theorem isomorphism to a four-dimensional vector space over a Euclidean field. The proof in Suppes (1959b) that invariance of relativistic distance along inertial paths is sufficient to derive the Lorentz transformations can be carried through over such a Euclidean field. Such a field can of course be denumerable, and consequently, the results are also interesting from the standpoint of the large philosophical literature on the problems of a metric or a measure in relativity. The intuitive reason that the proof can be carried through with just the apparatus of a Euclidean field available is that all the assumptions needed are macroscopic in character, and the algebraic methods of argument, although complicated in spots, are elementary, for example, familiar facts needed in the argument about affine spaces holding for affine spaces over Euclidean fields and not just over the field of the real numbers. From a pedagogical standpoint, this means that we should be able to teach the central mathematics of special relativity to students who have a good background in linear algebra, but who do not necessarily have any knowledge of the differential and integral calculus. However, I shall not push this point further here.

A third example is the algebra of physical quantities. By physical quantities I mean things such as lengths, times and masses; for instance 5 meters, 10 seconds and 15 grams are all examples of physical quantities. A detailed study of the algebra of such quantities is to be found in Chapter 10 of Krantz *et al.* (1971). Restricting ourselves only to square roots, for elementary purposes, we can easily give elementary axioms for physical quantities over an ordered Euclidean field. In these axioms, which are modifications of those given in Krantz *et al.* (1971), the set A is the set of physical quantities, in which fall the different dimensions of physical quantities that ordinarily occur in physics. We also include in the primitive notions the set A^+ for the positive physical quantities, a binary operation $*$ of multiplication of physical quantities, a unary operation $^{-1}$ for finding inverses and a unary operation $^{1/2}$ for finding square roots. Also, in stating the axioms the elements 0 and 1 of the given field are referred to. For more elaborate applications we will want to extend ourselves beyond a Euclidean field, but for elementary applications, this apparatus is sufficient. Consequently, I refer to the structures characterized in the axioms as elementary structures of physical quantities.

DEFINITION 5. *A structure* $\mathfrak{A} = \langle A, A^+, *, ^{-1}, ^{1/2}\rangle$ *is an* elementary structure of physical quantities (relative to an ordered Euclidean field $\mathcal{E}$) *iff, for all* x, y, z *in* A:

1. $x * y = y * x$;
2. $x * (y * z) = (x * y) * z$;
3. *If* $\alpha \in \mathcal{E}$ *then* $\alpha \in A$;
4. *If* $\alpha \in \mathcal{E}$ *and* $\alpha \in A^+$ *then* $\alpha \in \mathcal{E}^+$;
5. $0 * x = 0$;
6. $1 * x = 1$;
7. *If* $x \neq 0$ *then exactly one of* x *and* $(- - 1) * x$ *is in* A^+ ;
8. *If* x, y *are in* A^+ *then* $x * y$ *is in* A^+ ;
9. *If* $x \neq 0, x * x^{-1} = 1$;
10. $x^{1/2} * x^{1/2} = x$.

We may introduce the physical concept of dimension for such structures in the following way. If $x \neq 0$, the dimension of x is defined as:

$$[x] = \{\alpha * x | \alpha \in \mathcal{E}\}\,.$$

In other words, the dimension of x is just the set of physical quantities obtainable from x by multiplying x by a number, i.e., an element of the field $\mathcal{E}$. Of course, if we do not want to escalate the type of objects considered in elementary physics, we can introduce an equivalence relation instead of the set $[x]$. Physical quantities x and y have the same dimension, e.g., length, time, mass, force, etc., if there is a number α such that $\alpha * x = y$.

It is shown in Krantz *et al.* (1971) that an arbitrary structure of physical quantities can be represented as a multiplicative vector space over the rationals, or more exactly, a set of dimensions of such a structure is a multiplicative vector space over the rationals. Given this apparatus, we can then go on to elementary dimensional analysis, and more importantly, develop the elementary theory of the laws of similitude and exchange developed in Krantz *et al.* (1971).

I emphasize of course that I have given only a few samples of elementary physics in this brief discussion. It is, I think, worth finding out just exactly how much can be done within such a framework. One possibility, however, needs to be mentioned for enlarging the framework. If we want to make an exact connection with first-order logic on the one hand, and the usual background of real numbers on the other, it is natural to extend ourselves from Euclidean fields to real closed fields. Such fields are Euclidean, but they also have the property that every polynomial of an odd degree with coefficients in the field also has a zero in the field. A fundamental result of Tarski's decision procedure for elementary algebra and geometry is that any first-order sentence that holds for the field of real numbers also holds for real closed fields. By this extension, which takes us somewhat deeper into algebraic methods, we can get an exact correspondence between the two senses of elementary physics introduced at the beginning of this discussion.

21

ARISTOTLE'S CONCEPT OF MATTER AND ITS RELATION TO MODERN CONCEPTS OF MATTER

In this paper I want to analyze in some detail Aristotle's concept of matter. I do so not simply as a matter of historical scholarship, but in the interest of defending the correctness both scientifically and philosophically of what I would call the central doctrine. The elusiveness of Aristotle's detailed remarks on the concept of matter is notorious, and I shall not take it as my task to attempt to square my account with every passage that can be cited in the major works. I shall give references where they are obvious and appropriate. In some cases I shall assert features of his doctrine that are not properly documented in the text, but that I think are features of his concept of matter that are pretty generally accepted.

I also am not concerned to defend the details of all of his explicit beliefs. For example, what he has to say about the sun and the earth and the nature of circular motion is clearly false in detail. I am sure that if he had been presented modern astronomical evidence, especially astrophysical evidence about the swirling chaos of low density matter in outer space, he would have changed his views. Errors in detail of this

*Reprinted from *Synthese*, **28** (1974), 27-50.

kind seem to me to be of no importance. The basic doctrine, I argue, is correct. Moreover, I want to claim that it is correct in a strong sense: it can be used as a basis for interpreting the results of modern science. Defenders of Aristotle's concept of matter have been too defensive about the place of his concept in modern physics. I shall at the end of the paper attempt to put the case as strongly as I can for the correctness of Aristotle's view in the light of the best current knowledge about the nature of matter, as that term is ordinarily used by physicists. I am of course not suggesting that modern physicists talk about Aristotle or use in any obvious way an Aristotelian concept of matter. I do want to argue that they would often be better off if they did. Certain tendencies of research might indeed be improved if more heed were paid to Aristotle's doctrine than to the atomic theory we all tend so naturally and naively to accept. I think that it is very much a part of educated common sense at the present time to accept the building-block theory of matter in terms of atoms and molecules. We think of the spatial array of a molecule in terms of atoms, and we think of atoms as small planetary systems made up of simpler elements, such as electrons and protons. This building-block theory of matter is in detail obviously wrong. More importantly, it is conceptually wrong, and I want to argue that in spite of the importance for the history of science of the development of atomic views of matter in the nineteenth century and in the first part of this century, this aberration, like the aberration of universal determinism derived from classical particle mechanics, is mistaken.

I have organized my analysis in the following way. In the next section I state the central features of Aristotle's doctrine. After that, I compare this doctrine with modern scientific concepts of matter. Next I compare Aristotle's doctrine of matter with that of Descartes, Boscovich and Kant, in order to get a perspective on the philosophical thinking that parallels the development of modern science. In the final section I reexamine how Aristotle's concept of matter can be related to specific scientific theories of matter. I end with the strong claim that Aristotle's basic ideas are appropriate and proper for modern science.

1. CENTRAL FEATURES OF ARISTOTLE'S DOCTRINE

I have organized the features I want to emphasize under ten headings. I have not given under these headings a thorough account of Aristotle's doctrine of substance or his doctrine of motion, both of which are closely related to his concept of matter. I have tried to concentrate only on those features that are in my judgment most essential to his concept of matter.

(1) *Matter is the substratum of change. "For my definition of matter is just this—the primary substratum of each thing, from which it comes to be without qualification, and which persists in the result"* (*Physics*, 192a31; see also 190a15, 226a10, *Metaphysics*, 999b5, 104a32).

Matter as the substratum of change is perhaps the most characteristic aspect of Aristotle's doctrine of matter. It is important to keep in mind the relative concept as well as the ultimate one. In one sense the matter of the statue is the bronze from which it is made, yet the bronze itself is not ultimate matter but has itself various qualities such as heaviness and color. The rather delicate problem of how to talk about ultimate matter is discussed below in greater detail.

By putting Principle (1) first I also mean to emphasize the central physical character of Aristotle's concept of matter. Uses of the concept of matter as in talk about the matter of an argument or the matter of a geometrical line are taken to be clearly derivative and are not considered in any detail here.

(2) *A substance has both form and matter. The nature of a substance is complex. It is neither simply the form nor the matter* (*Physics*, 191a10, *Metaphysics*, 1043a15, and many other possible citations).

The distinction between substance and matter is critical for Aristotle. A substance is never pure matter. There are cases apparently in which the principle stated here is contravened in the other direction, however. It is possible to argue that according to his view the stars, for example, are substances that have no matter. I refer to this below, but for sensible substances of the kind that form the subject of the analysis of change, both form and matter are required.

(3) *Matter qua matter is purely potential and without attributes* (*Metaphysics*, 1029a19). *It is realized or 'actualized' only by some form. Consequently, matter as such cannot be properly defined* (*Metaphysics*, 1043b30, *Physics*, 194b8).

Principle (3) is fundamental for Aristotle's theory of matter. It is wrongheaded from his standpoint to ask of a substance what is its form and what is its ultimate matter and then to ask for properties of the matter. This view of matter seems contrary to that of contemporary physics with its talk about the quantity of matter or mass as an invariant property of matter. It must be realized that in talking about matter in this way physicists are not talking about matter in the way that Aristotle does. In abstract classical dynamics, for example, the only property of matter that is admitted is its mass, but even this admission is not consistent with

Aristotle's doctrine of matter as pure potentiality. The evident contradiction between these two ways of talking about matter does not mean that one is wrong and the other is correct—it means that the word *matter*, or its translation in various natural languages, is being used in more than one sense.

More importantly, Aristotle's own views divide naturally into statements about relative matter and statements about ultimate or prime matter. Contrary to Principle (3) it would be appropriate for Aristotle to ask about the properties of the relative matter of a bronze statue, for this is just to ask about the properties of bronze. It must also be conceded that many, if not most, of Aristotle's own remarks about matter are about relative matter not prime matter. The reasons for this should also be obvious. If we simply plunge from questions about the bronze statue to questions about its ultimate or prime matter, there is not much we can say that is appropriate, but this incongruity is no different from plunging into a modern analysis of the molecular structure of bronze.

The next two principles I want to discuss together.

(4) *There is no principle of individuation for matter qua matter.*

(5) *The principle of individuation for substances does not require sameness of matter for sameness of substance.*

Because matter qua matter is pure potentiality there are no attributes that can be used to characterize a principle of individuation for matter. (Note that in referring to matter qua matter here and earlier, I have in mind ultimate or prime matter, and thus an essentially equivalent formulation of Principle (4) is that there is no principle of individuation for prime matter.) On the other hand, we can use matter in differentiating some substances; for example, I can be holding two rocks and differentiate them by the fact that though their attributes seem to be the same they are composed of different matter. On the other hand, sameness of substance does not require sameness of matter. We talk about a physical body's being the same even though its matter may have changed; for example, a human body is both intaking and excreting substance, but we still speak of the identity of that human body through time.

There is a close parallel between the absence of a principle of individuation for matter and the problems of individuating points in space. One schematic way of describing the situation is in terms of observing in space the occurrence of some physical process or act. An example will suffice. Suppose we want to predict the height of the tide on the Pacific side of the Panama Canal two weeks from now at 0400 hr. Following a standard methodology we can represent the height of the tide measured by a vertical rod as a random variable with a given continuous probability distribution. For present purposes, it is useful to think of a 'question' with a yes-no answer as any interval on the measuring rod (technically we want not just intervals but any Borel set generated from intervals). If the tide falls within a given interval, the answer to the question posed by choosing that interval is yes, otherwise, no. Now suppose we consider two intervals, one being the closed interval [2m, 3m], and the other being the open interval (2m, 3m), the difference being that the first includes the two end points, 2 meters and 3 meters, and the other does not. Then our probability prediction will be the same for both intervals, and so will our claim that the observed tide falls within each interval. Our methodology of observation is not even in principle refined enough to discriminate between these two intervals, and this means our methodology does not permit us to individuate individual points, but only intervals of points. In this case points thus play a role analogous to that of prime matter.

Still another example of such a lack of a principle of individuation can be found in classical Zermelo-Fraenkel set theory with individuals. In a set theory of this sort, there is no satisfactory principle of individuation for individuals. It might seem appropriate to say that two individuals are identical if and only if they belong to exactly the same sets, but the identity of sets as formulated in the axiom of extensionality just depends upon two sets being identical if and only if they have the same members. So in such a set theory, as might be expected, we have no principles for asserting a principle of individuation for individuals. This of course is not surprising, because we have not built any structure that deals directly with individuals into the fundamental axioms. Indeed, with certain reservations, such a set theory with individuals constitutes a model for a fair number of the principles being stated in this section.

(6) *Substance has no contrary, but rather contraries like hot and cold are attributes of substance, and contraries can be attributes of the same substance at different times.*

For example, I may say that this pot is now hot, but it was cold when I started the fire a few moments ago. The pot itself does not have a contrary but each of its attributes can range from the attribute it now has to the contrary of this attribute, for example, from cold to hot. I shall have more to say about contraries after the statement of the next principle.

(7) *Only things or substances that change have matter* (*Metaphysics*, 1044b27). *Change is connected with the potentiality of opposites* (*Metaphysics*, 1050b26).

The contraries occupy a central role in Aristotle's theory of matter and of substance. What he has to say about these matters seems to me quite sensible, even though much of the talk on the surface seems very old-fashioned and far from talk of modern physics. The reason for this is not so much that the idea of contraries is now of no use but rather the kinds of examples he uses are not of great importance in physics itself; i.e., the concepts of hot and cold, for example, are replaced by the quantified concepts of heat and temperature, and more generally, the contraries represent a kind of qualitative theory of measurement that in most instances is replaced by a quantitative theory. In the subsequent analysis I shall not have much to say about the contraries, but it should be recognized that the doctrine of contraries is intimately related with the doctrine of matter as substratum, and it is not coherent to have a doctrine of matter as a substratum without something like a doctrine of contraries. The essential correctness of Aristotle's theory of contraries is represented by their continual use in ordinary talk. The scientific task has been not to establish the incorrectness of the contraries, but rather to provide a deeper-running quantified theory of the phenomena they describe.

I have avoided here the difficult problem of the generation and destruction of substances, and the analysis of contraries that is attached to the four elementary substances (e.g., in *Physics*, 189b). The last chapter of Book I of the *Physics* does seem to yield a relatively straightforward argument for the conservation of prime matter, but since an explicit conservation law seems contrary to the spirit of Aristotle's view, I have omitted a separate statement of such a principle. It does seem needed in any attempt to make explicit the theory of generation and destruction of primary substances.

(8) *The matter of a body or substance is not the place of the body or substance, and is not therefore that which contains the body or substance. Put another way, the matter of a body or substance cannot be identified as the container of that body or substance* (*Physics*, 209b22 and 211b30).

This principle is a clear enunciation that matter is not space and a container theory of matter is not part of Aristotle's doctrine. I highlight it here because it is, under one interpretation, in direct contradiction with Descartes's theory of matter as extension, which I discuss in Section 3.

(9) *The void does not exist as a separate thing or substance. The most that can be said is that "the matter of the heavy and the light, qua matter of them, would be the void "* (*Physics*, 217b22).

This principle I include to separate Aristotle from the classical atomistic tradition and, for example, from the theory of matter advocated by Boscovich. More importantly, the idea of empty space has been central to atomic doctrines, both ancient and modern, but it is also true that since the discovery that light and other electromagnetic phenomena are propagated with finite velocity there has been little tendency to accept the void as a serious physical concept. It is also part of this principle that Aristotle does not accept that matter is made up of indivisible homogeneous simple elements that exist in a void. In other words, the atomic theory of matter is inconsistent with Aristotle's.

(10) *The sun and stars have no matter; their motion does not involve the potentiality of opposites; circular motion has no contrary (On the Heavens*, 270a12; *Metaphysics*, 1050b22).

As indicated earlier, this principle of Aristotle seems mistaken, but I do not take the mistake to be a serious one.[1] On the basis of modern evidence I am sure it is the one principle of the ten that he would have changed. It seems to me that the remaining ten can stand essentially unaltered. I do not mean that there are no other statements of Aristotle about matter that need correction, but of the features that I consider characteristic of his doctrine, it is only this last that seems to me to be clearly and unequivocally in error. The error is in a major application of the general theory, not in the general theory itself.

2. MODERN SCIENTIFIC CONCEPTS OF MATTER

Before examining in more detail Aristotle's concept of matter it may be of some value to relate it in a general way to modern concepts of matter. There are two great traditions to be examined. One is the philosophical tradition and the other is the scientific. In the earlier period, of course, these traditions were not sharply separated. I take as prime examples of the philosophical tradition Descartes, Boscovich and Kant. Descartes and Boscovich both thought of their contributions as being part of science as well, insofar as there was any clear separation between philosophy and science in the seventeenth century and in the framework of the eighteenth century within which Boscovich operated. Kant clearly separated his own contribution, especially the metaphysical foundations of natural science as set forth in the work of that title. The scientific tradition, on the other hand, is associated with the development of atomic theories of matter

[1] Aquinas, in the *Treatise on Separate Substances*, takes the firm position that the heavenly bodies have both form and matter.

in the nineteenth century and the deep development of particle physics and quantum mechanics in the twentieth century. It is characteristic of the scientific tradition that it is difficult to find explicit and categorical answers to the question, what is matter? The view of matter that may be inferred from the scientific tradition is, however, often fairly obvious. It would take us too far afield to try to examine the history of that development in detail. It may be useful to say something about it before turning back to the philosophical tradition.

Certainly one conclusion that can be drawn is that even implicitly there seems to be nothing close to Aristotle's concept of prime matter in the scientific developments since the end of the eighteenth century. Much of the initial thrust was to revive and actually develop a very viable theory of atoms, a theory that certainly is closer to the ideas of Democritus and Epicurus than of Aristotle. In the latter part of the nineteenth century electromagnetic theory and the experiments connected with the development of special relativity create an atmosphere that is more congenial to Aristotle's ideas, but the remoteness of these developments from Aristotle is exemplified by the fact that in E. T. Whittaker's exhaustive *History of the Theories of Aether and Electricity* (1910) there is no mention of Aristotle whatsoever. Of course, it is possible to attribute this to ignorance on the part of the scientists responsible for the theories of the ether and electricity, and it is even possible to attempt to claim that the concept of the ether is itself closely related to Aristotle's concept of prime matter. However, a little reflection indicates that this is a futile hope. Certainly it is completely inconsistent with Aristotle's characterization of matter to attempt to build the kind of mechanical model of the ether for which Lord Kelvin and Maxwell are famous. The definite attribution of mechanical and electrical properties to the ether is inconsistent with Aristotle's conception of prime matter as pure potentiality. In fact, the main thrust of the nineteenth-century models of the ether was to apply the relatively deep mathematical and conceptual developments of the mechanics of fluids to the construction of mechanical models of the ether, with the addition possibly of separate and independent electromagnetic properties.

It must also be recognized that from the end of the nineteenth century and through the development of quantum mechanics, the acceptance of the electron as a fundamental particle of an indivisible and fixed character with definite mass and charge is very much in the spirit of Democritus and atomism, rather than in the spirit of Aristotle's physics, just as was the case a hundred years earlier in the development of the atomic theory of matter. I know of no serious discussion that relates Aristotle's concept of matter to the theory of fundamental particles running from, say, 1890 to 1930.

There are two further remarks I want to make about Aristotle's concept of matter in connection with modern scientific theories of matter. The first concerns axiomatic foundations of modern theories. It might be thought that even though the formulations of theories of matter by physicists do not invoke a concept at all close to Aristotle's this is simply due to their leaving implicit major assumptions. It is well known, for example, that foundational discussions of physics do not in general satisfy the most rudimentary mathematical standards of explicitness from an axiomatic standpoint. It might be felt that an explicit axiomatic theory of mechanics or electromagnetic theory would bring out closer connections between Aristotle's theory of matter and contemporary scientific theories. The contrary seems to be the case.

If we consider, for example, axiomatizations of particle mechanics, we take as undefined or primitive the set of particles but immediately attribute properties to these particles, especially mass. As we move on to more complicated objects like rigid bodies we attribute additional fixed properties like those of moment of inertia. When we turn to electromagnetic theory we encounter attribution of charge or, in the case of electromagnetic fields, measures of intensity of the field that are meant to be in principle observable. Nowhere in such discussions is there a hint of something corresponding to Aristotle's distinction between form and matter.

There is one possible exception to these remarks; it is the case of classical continuum mechanics, to which I return in Section 4.

The second remark concerns the apparent instability of current concepts about elementary particles and the general chaos of theory in high energy physics. When it was thought that there were a few fundamental particles out of which everything else in the universe was composed and that these particles were themselves indestructible and in some clear sense elementary simples, then the atomic theory of matter seemed to have won the battle, even if the elementary particles did not possess all the properties we expect of macroscopic bodies. Research in physics of the last couple of decades has shown that this picture is not at all the correct one. The number of particles has been shown to be very large, and there is now some skepticism that any simple account in terms of a few fundamental particles will ever be made to work. Certainly it would seem that the present situation in high energy physics is much more congenial to an Aristotelian theory of matter than the situation that obtained even 30 years ago. I shall have something more to say about these matters in Section 4.

3. SOME COMPARATIVE PHILOSOPHICAL CONCEPTS OF MATTER: DESCARTES, BOSCOVICH AND KANT

If philosophical developments closely followed the scientific developments just sketched, then little sign of Aristotle's influence on modern philosophical concepts of matter would be expected to be found. If we look at the most influential concept of matter in the seventeenth century, the century that ushers in modern science, then all traces of Aristotle seem to have disappeared. I refer of course to Descartes's concept of matter. This also seems to be true when we look at Boscovich's influential views in the eighteenth century, but the situation is quite different when we come to Kant.

To provide a broader framework for analyzing Aristotle's concept of matter, I shall briefly examine the concept of matter advanced by each of these three philosophers.

The most systematic exposition of Descartes's physical theory is to be found in his *Principia Philosophiae* (1644). Part II of this treatise is concerned with the general principles of material things that can be known clearly and distinctly. In this part are established a large number of general propositions concerning the nature of matter, the existence of atoms, the laws of motion, etc. As is well known, Descartes attempts to describe and explain the physical world in terms of nothing but extension and motion. The fundamental characteristic of matter or body is extension (I, Art. 53, II, Art. 4).[2] This property of extension is the only clear and distinct idea of body that we can have (I, Art. 54, Art. 63, II, Art. 1). On the other hand, matter qua extension is obviously undifferentiated, so there is a difficulty to explain the variety and diversity of bodies. Descartes's answer is given in terms of motion, "All the variation in matter, or diversity in its forms, depends on motion" (II, Art. 23). The only kind of motion admitted is of course local motion, and the proper definition of motion is "the transference of one part of matter or one body from the vicinity of those bodies that are in immediate contact with it, and which we regard in repose, into the vicinity of others" (II, Art. 25).

Descartes gives a succinct summary of his theory in the following passage (IV, Art. 203).

> Having considered in general all the clear and distinct notions that can be in our understanding concerning material things, and not having found any of these other than those of figure, size, and motion, and the rules according to which these things can be diversified by one another, which rules are the

[2] References refer to Parts and Articles of Descartes's *Principia*.

> principles of geometry and mechanics I judged that all the knowledge that men could have of nature had necessarily to be derived from this only; because all the other notions that we have of sensible things being confused and obscure, cannot serve to give us knowledge of anything outside us.

There is a great deal of additional detail in the *Principia*, but it is inordinately tedious to read, and we can well believe Gassendi's remark that he knew no one who had read the work in its entirety. The features of Descartes's theory that I have presented here are sufficient to recognize its conceptual inadequacy. Descartes's reduction of the concept of body to that of geometrical solid and his use of a purely relational definition of motion made it impossible for him to give a consistent extension of these ideas from kinematics to dynamics. His own account of forces is a shambles and is simply a reflection of the inadequacy of Descartes's ideas for the development of any serious conceptual framework for physics.

The greater subtlety and empirical adequacy of Aristotle's ideas are evident, and it may seem something of a puzzle to understand why Descartes's ideas had the enormous influence they did in the seventeenth century. (This influence has been well documented in the classic work of Mouy (1934).)

Of course, the simplicity and surface clarity of Descartes's prose is enormously appealing in contrast to the Proustian quality of the commentators on Aristotle. In any case, the change from one set of philosophical ideas to another is not a process that we understand very well or have as yet studied with any thoroughness. It is still astounding to find Descartes taken so seriously, but not nearly as astounding as other philosophical examples that could easily be cited.

Boscovich, operating almost a hundred years later, adopted a methodology very similar to Descartes's but in many respects stood Descartes's theory on its head, though he remained as far from Aristotle as did Descartes. The analysis of his concept of matter I give here is restricted to his major work, the *Theoria Philosophiae Naturalis*, which was first published in Vienna in 1758 and then in a revised form in Venice in 1763. (References are to articles of this work.)

In the first six articles, Boscovich states what he has in common with Newton and Leibniz, and how his own theory differs from theirs. His nonextended points are similar to Leibniz's monads, and the mutual forces acting between them are extensions of Newton's ideas about forces. He differs from Leibniz in making his points homogeneous and denying the principle of indiscernibles and the doctrine of sufficient reason. He differs from Newton, he says, by using repulsive forces as well as attractive ones.

Boscovich thinks that his greatest achievement was to improve on Newton and reduce phenomena to one principle, his single law of forces. He felt that his chief intellectual debts were to Leibniz and Newton, and in his own mind his relation to the Cartesians is primarily negative. Aristotle plays little part in the explicit discussion of his theory.

The kernel of Boscovich's theory of matter is easily summarized. The matter of the universe is composed of a finite number of nonextended points: attractive and repulsive forces, which are a function of distance only, act between these points according to a single law of forces. All the observed phenomena of nature are to be explained solely in terms of the distribution and motion of these points and the forces acting between them. In his own picturesque phrase, "matter is interspersed in a vacuum and floats in it." (Art. 7)

The principle of the nonextension of matter and the law of forces are the two fundamental hypotheses of Boscovich's theory, but they are not presented as axioms from which verifiable consequences are deduced. Instead, a plausible derivation of them from the more familiar and generally accepted laws of impenetrability and continuity is given. I shall not enter into these details here, but Boscovich's arguments provide indirectly an excellent critique of the Cartesian ideas and bring out inconsistencies in the Cartesian notions.

Boscovich reaches four main conclusions about the primary elements of matter. The first one is that the parts of matter are not contiguous, and the second is that the primary elements are simple, for if they were composite, the indefinitely large repulsive forces would drive the pieces asunder. Boscovich states his view very clearly:

> Now, because the repulsive force is indefinitely increased when the distances are indefinitely diminished, it is quite easy to see clearly that no part of matter can be contiguous to any other part; for the repulsive force would at once separate one from the other. Therefore it necessarily follows that the primary elements of matter are perfectly simple, and that they are not composed of any parts contiguous to one another. This is an immediate and necessary deduction from the constitution of the forces, which are repulsive at very small distances and increase indefinitely. (Art. 81)

The third conclusion about the primary elements of matter is more uniquely Boscovich's own than the first two. It is that the elements are nonextended. The direct argument runs as follows. Since the elements are simple, they cannot have extension of the ordinary sort, but the question arises: can they have what the Scholastics called "virtual extension"?

Virtual extension compared with actual extension can for our purposes probably best be understood by giving an example or two. God, who is perfectly simple, is yet everywhere. In the same way, some have argued that the soul is simple and yet (virtually) extended throughout the whole body.[3] Boscovich is willing to admit that it is metaphysically possible that the primary elements of matter possess such virtual extension, that is, that it cannot be proved on metaphysical grounds that they do not (Art. 83). However, on empirical grounds he argues it can be shown that they do not possess virtual extension. If virtual extension were a property of bodies of sensible size, we would be able to observe it. No such observations have ever been made. "Further, this property by its very nature is of the sort for which it is equally probable that it happens in magnitudes that we can detect by the senses and in magnitudes which are below the limits of our senses." Thus, since it is not observed in the one case, we may infer by induction that it does not occur for the primary elements of matter that cannot be directly observed (Art. 84). This discussion of virtual extension is one of the less satisfactory aspects of Boscovich's analysis. It is simply part of his argument to stand fast on the view that the primary elements of matter are nonextended. A good many additional arguments about nonextension are given, especially in Articles 88–90.

The fourth conclusion about the primary elements of matter is that they are homogeneous. Boscovich offers several arguments in support of this conclusion. One argument depends on the law of forces. The curve of forces is the same in its two asymptotic branches for all elements, since all are equally impenetrable and subject to gravitational action. Now there are infinitely many more curves "which, when they differ in the remaining parts, also differ to the greatest extent in the extremes, than there are curves, which agree so closely only in these extremes" (Art. 92). Hence, Boscovich asserts, it is infinitely more probable that the curves agree in all their parts than that they differ between their identical extremes. (Another and rather similar argument is adduced from the similarity of bodies (Art. 96, Art. 97).) The Leibnizian objections to homogeneity on the grounds of the principles of sufficient reason and indiscernibles are rejected with supporting arguments. A vivid analogy using books, letters and dots is used to complete the arguments for this fourth conclusion. Assume a method of printing that prints each letter as a dense series of small, similar black dots (rather like many modern computer printers).

[3] This Scholastic notion of virtualness is hard to give empirical content. Typical examples of another sort help illustrate its meaning: a pentagon virtually contains a quadrangle and a quadrangle virtually contains a triangle; a man is virtually an animal, and an animal is virtually a plant.

From the letters of the alphabet all words used in books are formed. Thus the enormous diversity actually to be found in books can be accounted for by the distribution of many similar black dots. The analogy runs this way. Books correspond to gross bodies. The different substances found by chemical analysis correspond to words. Further chemical analysis discloses a few fundamental particles that correspond to the letters. And finally the dots composing the letters correspond to the simple, homogeneous primary elements of matter (Art. 98, Art. 99).

It seems fair to say that Boscovich's theory represents the thorough working out of the ancient atomistic tradition, and he represents the carrying of this tradition to its finest point. He has, like Descartes, the virtue of offering an extraordinarily simple and clear theory. It is unfortunate that it just turns out to be so thoroughly unworkable and inadequate. It seems to me that in many ways Boscovich's theory represents the fantasies of many physicists, who would like to find that matter is made up of ultimate simples that have exactly the properties predicated of them by Boscovich.

We can see that Boscovich is the opposite of Descartes in affirming that matter is nonextended and that empty space is everywhere, but in the simplicity of his basic conceptions there lies strong affinity to Descartes. Given the great simplicity of Descartes's or Boscovich's ideas, it might seem that there would be little hope of reviving the subtler and more difficult Aristotelian ideas, even if the ideas of Descartes and Boscovich turned out to be wholly inadequate in providing a framework for actual physics.

Kant provides a counterexample. His ideas about substance are much closer to Aristotle's than to Descartes's or Boscovich's. Aristotle's basic argument about substratum, i.e., there must always be something underlying that which is in the process of becoming, is essentially Kant's argument for the existence of substance. It will be worthwhile to look at some of the details.

I shall mainly deal with Kant's views on the nature of matter as set forth in the *Metaphysical Foundations of Natural Science*, but I shall also make reference to significant passages about substance in *The Critique of Pure Reason*. Kant's use of the categories to find the specific determinations of matter is another Aristotelian aspect of his theory of matter. There are some difficult problems about the relationship between the concepts of matter and motion for Kant, and I do not want to enter into these problems in detail here. I have discussed them elsewhere (Suppes, 1967). For the purposes of our discussion here I think we may claim that Kant held that the concept of matter includes the concept of an object of the external sense and that this latter concept includes the concept of motion. Whether this is exactly the correct story, Kant does assert unequivocally that we may reduce all proper natural science to a pure or applied theory of motion. It is then as the doctrine or theory of motion (Bewegungslehre) that the metaphysical foundations of natural science are brought under the four divisions of the table of the categories. In the first division, matter is considered purely according to its *quantity* of motion, abstracted from all its qualities. This gives us the theory of phoronomy or kinematics. In the second division, motion is considered as belonging to the *quality* of matter, "unter dem Namen einer urspruenglich bewegenden Kraft". This yields dynamics. The third division is mechanics; here, motion as quality is considered in *relation* to other reciprocal motions, or, more exactly, matter with this dynamical quality of possessing an original moving force is considered in reciprocal motion. In the fourth division, entitled phenomenology, matter in motion or at rest is considered according to its *modality*; that is, whether in its determination as a phenomenon of the external sense it is determined as possible, real or necessary.

If we left matters at this level of generality, it might seem that there was an enormous similarity between Kant's and Aristotle's theory of matter. However, the special role that Kant assigned to fundamental forces of repulsion and attraction moved the development of his ideas away from a purely Aristotelian framework. Kant emphasizes that the fundamental forces of repulsion and attraction cannot themselves be constructed; their possibility cannot be demonstrated. These fundamental forces are not derived from experience, nor can they be mathematically constructed from other concepts, which would be necessary to demonstrate their possibility. They are jointly the ultimate ground for the possibility of matter. If one asks why matter fills its space by these original forces, the only answer is that they are necessary conditions for the construction of the concept of matter. Reason can do no more than reduce the diverse forces appearing in nature to these two fundamental ones, "beyond which our reason cannot go".

If the fundamental forces cannot themselves be comprehended or explained, if they are each the source of an ultimate explanatory principle, and if the concept of them is used to construct the concept of matter, then the delicate problem arises: of what are these forces predicated? Is it a vicious circle to say they are forces *of* matter? Would it be more nearly correct to say that these forces *are* matter? This is not the same as asking for an explanation of the forces. Rather, accepting them as ultimate, we are asking the different question: to what do they belong, if anything? Boscovich answered this question by making forces ultimate in nature, but retaining as carriers of the forces a finite set of points of singularity. For Boscovich, forces are predicated of these points, which for him solves the question that we are now asking Kant. Kant eliminates all points of singularity in space that might serve as ultimate subjects of the forces. Empty space cannot be an object of experience, and every part, i.e., every point, of filled space possesses forces of attraction and repulsion. Now it is tempting to say that in abolishing all points of singularity and predicating forces of every point of space that can be experienced, Kant has unequivocally adopted a complete dynamical theory of matter and has asserted that forces are matter. There are passages in the Dynamics that lend definite support to this view. For instance, the General Remark on the Dynamics begins: "The universal principle of the dynamics of material nature is: that all reality of the objects of the external sense, which is not mere determination of space (place, extension and figure), must be regarded as moving force...." However, there does not seem to be a fully adequate case for this view. The discussion of substance in the *Critique* forms one of the chief difficulties for such an interpretation. The first analogy of experience states the principle of the permanence of substance. This analogy is the rule corresponding to the category of inherence and subsistence. The principle states that in all changes of phenomena, substance is permanent and is neither decreased nor increased (*Critique*, B224). Substance is simply the substratum of all determinations in time, i.e., of all changing phenomena. Kant's argument is that the bare succession of phenomena must have a permanent substratum as a necessary condition, for this substratum is "the condition of the possibility of all synthetical unity of perceptions, that is, of experience" (*Critique*, A183, B226-27). Without this substratum, the manifold of phenomena given in time could not be determined according to any rules, and could not be connected as objects enduring in time.

The second analogy of experience, which corresponds to the category of causality, is that all changes take place according to the law of causality. For the moment, the important point of this is that changes must be changes in the determinations or states of the permanent substance, one state following another according to a given rule. The permanent substance provides the ground for the connection of successive states; in fact, if substance were created or destroyed, the universality of the law of causality would be violated (*Critique*, B232-33).

But what is the empirical criterion of substance? "Action...is a sufficient empirical criterion to prove substantiality, nor is it necessary that I should first establish its permanency by means of compared perceptions, which indeed would hardly be possible in this way, at least with that completeness which is required by the magnitude and strict universality of the concept" (*Critique*, A205, B250-51). Action directly implies the relation of the subject of causality (substance) to the effect. But for action there is needed the permanent substratum, for "actions are always the first ground of all change of phenomena, and cannot exist therefore in a subject that itself changes, because in that case other actions and another subject would be required to determine that change" (*Critique*, A205, B250). Actions, forces, cannot subsist by themselves but must be determinations of a permanent substratum. On the other hand, Kant says, substance "appearing in space," that is, matter, can only be known to us through the two fundamental forces of attraction and repulsion. Other properties of matter are unknown to us (*Critique*, A265, B321).

Without going further into the systematic discussion of substance in the *Critique*, I believe we may now answer the question we asked about the fundamental forces. Matter, as spatial substance, as the ultimate subject of the science of physics, is not simply the two fundamental forces. It is true that the concepts of these two forces are precisely those that permit us to construct the concept of matter, i.e., represent it in intuition; and simply as an object of intuition, matter is equivalent with them. However, matter as substance is also the permanent substratum of all spatial phenomena. The fundamental forces are not this permanent substratum, but rather it is "the amount of the fundamental forces" possessed by a given part of this substratum that determines its particular state. The mathematician or physicist, dealing as he does only with pure or empirical intuitions, might successfully equate the fundamental forces and matter; but the philosopher, probing at the foundations of the data of intuition, knows that the fundamental forces are not the ultimate subject in space, but are the specific determinations of that subject (the permanent substratum). And this conclusion is supported in the third division of the *Metaphysical Foundations*, where Kant specifically states that the quantity of substance in a matter, that is, the quantity of the permanent substratum, is not a function of the amount of the fundamental forces in that matter, but must be estimated mechanically, that is, by the amount of its motion.

It seems to me that this discussion of force and matter in Kant—the delicate effort he makes to assign a fundamental place to force, and yet not eliminate an independent concept of matter—is still pertinent today. It is particularly relevant to the tangled problems of thinking about force, matter and energy, in any conceptually clear way, in the context of contemporary nuclear physics. I do not mean to suggest that detailed answers for today's puzzles are to be found in reading Kant. I do think that some of the too-simple models we associate with the Cartesian and Newtonian tradition would be more easily rejected as inadequate on general philosophical grounds if we took seriously Kant's careful and discriminating analysis.

Kant's dynamical forces are certainly not a part of Aristotle's theory of matter, but the discussion of substance as substratum is very much in the Aristotelian spirit, and shows clearly enough that Aristotle's fundamental ideas were restored to the mainstream of philosophical discussions of matter by Kant.

4. SCIENTIFIC RECONSTRUCTION OF ARISTOTLE'S CONCEPT OF MATTER

As I promised earlier, I want to end by making a case for the scientific relevance of Aristotle's concept of matter to contemporary physics. There are three directions of attack I think can be successfully taken. One is in terms of the modern evidence on elementary particles, the second concerns modern work on the foundations of classical mechanics and the theory of bodies in classical mechanics, and the third is the attitude toward the use of random variables in probability theory. I shall only discuss the first two lines of attack in this paper, and reserve the random-variable analysis for another occasion.

As the atomic theory of matter became a workable empirical theory at the beginning of the nineteenth century, it looked certain at that time that the ancient atomic theories of matter were the conceptually correct ones, and all that was left was to work out the details of the interactions of the fundamental atomic parts of matter.

By the end of the nineteenth century it was recognized that atoms have structure, and aspects of this structure were clearly identified. The concept of a nucleus with electrons "in orbit" around the nucleus was developed, and everything seemed once again quite satisfactory. The atom was thought of on the lines of a small-scale solar system, and the fundamental particles were now not atoms, as atoms had been identified earlier in the century, but electrons and protons. It also seemed clear that these elementary particles had fundamental constant properties, for example, a fixed mass (rest mass as the theory of relativity developed), a fixed charge and a negligible but definite size.

As quantum mechanics developed and the many experimental anomalies in the classical picture of the structure of the atom were identified, it became apparent that the particles that make up an atom were not simply little balls bounding around in a small-scale world very much like the one we observe. The properties were peculiar and the theory was tantalizingly elusive. It was also recognized that matter was not indestructible, contrary to ancient ideas of an atomic sort, but that it could be converted into energy. Still, the case for the atomic theory in some form seemed strong, and most physicists probably felt that some version of the atomic theory was basically the correct theory of how the universe was put together. Even if electromagnetic and possibly gravitational fields were admitted, the atomic theory together with some kind of theory of the ether seemed to create a plausible picture.

The pursuit of particles continued and as the energy levels became higher it became apparent that the world is full of particles that are continually undergoing processes of generation and corruption, as Aristotle would put it. Methods for observing this generation and corruption were brought to a fine point by bubble-chamber apparatus and other related methods.

It does not seem to me necessary to fill in the details of this picture in order to describe in qualitative terms how Aristotle's theory of matter fits in. From Aristotle's standpoint, the search on the basis of the evidence available for fundamental building blocks is a clear mistake. The empirical evidence from macroscopic bodies and also from high energy particles is that the forms of matter continually change. There is no reason to think that there is a spatial buildup of electrons, for example, from some more elementary objects. The collisions of electrons and other particles to produce new particles as observed, for example, in cloud-chamber and other experiments is simply good Aristotelian evidence of the change of form of matter. The cloud-chamber data especially support Aristotle's definition of matter. As we observe change there must be a substratum underlying that which is changing. What is the substratum underlying the conversion of particles into other particles, or the conversion of particles into energy? The answer seems to me clear. We can adopt an Aristotelian theory of matter as pure potentiality. The search for elementary particles that are simple and homogeneous and that are the building blocks in some spatial sense of the remaining elements of the universe is a mistake. There is a continual conversion of the forms of matter into each other; there is no reason to think that one form is more fundamental than another. The proper search at a theoretical level is for the laws that describe these changes of form, and not for the identification of elementary particles that are in some fundamental and ultimate sense simple and homogeneous.

In summary, the case seems good for Aristotle's theory of matter providing an excellent way of looking at the phenomena of high energy physics as well as at the macroscopic kind of phenomena Aristotle himself had available. I do not mean to suggest that we can pull any detailed wide scientific laws from Aristotle. What is valuable in his concept is its wide applicability as a way of thinking about physical phenomena.

Kant was right in his criticism of the Cartesian mechanical method, but he was wrong in a way that Aristotle was not in attempting too simple an account of the fundamental forces of nature.

This sketch I have given of the way in which Aristotle's theory of matter can be used to provide a sound interpretation of the proliferation of particles and processes in high energy physics needs of course to be spelled out in greater detail, but it seems to me that its essential soundness is easy to recognize in spite of the broadness of the strokes I have used.

Classical mechanics of bodies. It will be useful to end with a more detailed and technical treatment. The reader who is unfamiliar with the manifold problems encountered in the exact statement of the foundations of classical mechanics may think that there is little new to be said about this subject, and that there is scarcely a proper place for the Aristotelian concept of matter. The point I wish to emphasize is that the mathematical and conceptual difficulties of classical mechanics are severe. We are still far from a completely satisfactory general theory. There is, on the other hand, a very substantial gain in clarity and understanding that has taken place in the last decade or two, especially due to the work of Walter Noll, Clifford Truesdell and others. It is fair to say that there has been a renaissance of classical mechanics.

I shall end with a sketch of Noll's (1959) axioms for bodies and their kinematic motion. I shall omit some of the technical mathematical details required for formulating smoothness conditions.

DEFINITION. *A* body *is a set B endowed with a structure defined by a set Φ of mappings of B into a three-dimensional Euclidean space E, and a real-valued set function M defined for all Borel subsets of B, subject to the following axioms:*

(1) *Every mapping ϕ in Φ is one to one.*

(2) *For each ϕ in Φ the image of B under ϕ is a region in the space E, a region being defined as a compact set with smooth boundaries.*

(3) *The mass function m is a nonnegative measure.*

(4) *For each ϕ in Φ the measure induced by m on the region that is the image of B under ϕ is a mass-density function that is positive and bounded.*

Following Noll, we may refer to the elements of B as the *particles* of the body, the mappings ϕ in Φ are the *configurations* of the body. If a is in B, and ϕ is a configuration, then $\phi(a)$ is the *position* of the particle a in the configuration ϕ. The set function m is the mass distribution of the body and the density under the mapping ρ_ϕ is the mass density of B in the configuration ϕ.

A motion of a body B is a one-parameter family $\{\theta_t\}$ of configurations θ_t in Φ of B such that the derivative $(d/dt)\theta_t(a)$ exists for all a in B and all times t. The derivative is a continuous function of a and t jointly, and is a smooth function of a. Moreover, the second derivative also exists and is piecewise continuous in a and t jointly. (The first derivative is the velocity of the particle a at time t, and the second derivative is the acceleration of a at time t.)

From these definitions, we can go on to develop a comprehensive though not completely adequate theory of bodies in classical mechanics, where bodies are not just rigid bodies but the sorts of configurations to be encountered in continuum mechanics. They would be covered grammatically, for example, by mass nouns. To complete the development of the present theory we need to add appropriate definitions of body forces and contact forces, and to define the general concept of a dynamical process, but these matters will be omitted here.

Instead, I want to turn to some closing remarks about how Aristotle's ideas of matter may be fitted into this framework, and also to indicate what some of the difficulties are. At one level, the situation seems clear. We simply identify the prime matter of the body as the set B without structure. The introduction of structure corresponds to Aristotle's introduction of forms. The configurations provide the geometrical shapes of the body through time, and the mass function the distribution of density through time. It should also be evident that the particles talked about here are of course not atomic particles or elementary particles in the sense of physics. These are idealized particles that make up a continuum and for this reason they come reasonably close to Aristotle's idea of matter, even though we do use the particles themselves as arguments of functions and thereby in one sense endow them with attributes in a way that he would consider incorrect. I think however that we can take the attitude that the configurations change and therefore we are not endowing a particle, as such, with an attribute, but this is the way of introducing forms.

The important point is that the set B is not like a set of persons or a set of individuals with structure, as for example a set of bronze statues, but is indeed a set that, taken without structure, seems very close to what Aristotle had in mind. I shall close by modifying the first part of the definition of bodies.

> *A body is matter endowed with a structure. We represent the matter by an abstract set B and the structure by a set ϕ of mappings of B into a three-dimensional Euclidean point space*

E and a real-valued set function m defined on the Borel subsets of B, subject to the axioms stated above.

This definition satisfies fairly well the ten Aristotelian 'principles' of matter stated in Section 2, but, of course, Principle (10) concerning the heavenly bodies does not really apply. I say "satisfies fairly well", because the principles are not stated in a sufficiently formal manner to make satisfaction of them a completely objective affair. For example, Principle (4) concerning individuation of matter qua matter needs detailed analysis, but very much along the lines already given in Section 2.

Reconstruction of the concept of matter along the lines of the formal definition just given does not, however, do justice to what is probably the most important insight of Aristotle concerning the concept of matter. This is the relative sense in which bricks, for example, are the matter of a house, and clay, the matter of bricks. As things have turned out the relative sense of matter has not become a fundamental concept in modern science, but its practical importance in science as well as in ordinary affairs is easily recognized. Biologists, for instance, almost always use such a relative concept of matter, even if it is not so labeled. A more systematic and explicit analysis of the way in which the relative concept of matter can or does enter in various modern theories of science would seem desirable.

22

POPPER'S ANALYSIS OF PROBABILITY IN QUANTUM MECHANICS

Since the early 1930s, Popper has been publishing articles about the foundations of quantum mechanics. and he has had many useful things to say in a number of books and articles. With some hesitation, I have narrowed the scope of what I shall discuss to two topics: the propensity interpretation of probability and quantum mechanics as a statistical theory. Throughout his writings, but especially in *The Logic of Scientific Discovery*, Popper has a great deal to say also about measurement in quantum mechanics. I shall not discuss these matters except as they bear upon his conception of the role of probability in quantum mechanics.

1. PROPENSITY INTERPRETATION OF PROBABILITY

Popper has spent considerable effort in elaborating and defending his propensity interpretation of probability, and the term itself has become a widely used one for a certain conception about probability. More than anyone else, he has been responsible for making this view of the foundations of probability well known, and he says in several places that he was

*Reprinted from *The philosophy of Karl Popper, Volume 2* (ed. by P. A. Schilpp), 1974, pp. 760–774. La Salle, IL: Open Court.

led to the propensity interpretation as part of his reflections on the role of probability in quantum mechanics.

I shall concentrate on comparing the propensity interpretation with the other major views of the foundations of probability: the classical Laplacean definition, the relative-frequency interpretation, and the subjective interpretation. Only after a fairly detailed scrutiny of the propensity interpretation as an alternative to one of these three positions will I have anything to say about the particular case of quantum mechanics. I shall not discuss confirmation or corroboration as an additional way of looking at the foundations of probability. I think a case can be made for considering confirmation theory as a fourth alternative; but Popper has rather different views of this matter, and some rather special views of his own about corroboration. I believe the issues surrounding the propensity interpretation can be discussed by restricting the frame of reference to the three major views mentioned.

To begin with, I think Popper has brought to the surface some intuitions that many of us share about the foundations of probability. He has had the insight to recognize that there is something misleading about each of the main classical views of the foundations of probability in at least some applications. The propensity viewpoint or what we might also call the dispositional viewpoint toward probability is very appealing, not only when we deal with the physics of atomic and subatomic particles, but also with many straightforward applications in medicine, psychology, sociology, etc. Moreover, Popper has properly emphasized that we cannot simply think of the propensity or disposition as inhering in the object independent of the circumstances surrounding the object. In other words, propensity, as Popper remarks, like the concept of force "is a relational concept." [1]

Although I am sympathetic with these intuitions and with the insight that Popper has verbalized for all of us, a central issue about the propensity interpretation dominates all other issues for me, and I would like to concentrate on it in this discussion. In broad terms, the issue is that of characterizing the explicit meaning of the propensity interpretation. To ask for the meaning of the interpretation without making any more definite statement of the methodology to be used or the approach to be taken is to ask in philosophers' jargon the ordinary question, What is the propensity interpretation? However, in the case of probability, a strong intellectual tradition of analysis can be used.

The matter can be formulated this way. With an important exception in quantum mechanics to be discussed below but not an exception

[1] K. R. Popper (1959). Hereinafter cited as *PI*.

that affects the conceptual point being made at this stage, the scientific applications of probability all rest on the acceptance of Kolmogorov's axiomatization, and the formal properties of probability that flow from that axiomatization. On occasion, Popper has indicated some reservations about that axiomatization (PI, p. 40), but those reservations are minor and not a serious issue here. The mathematical applications of probability in sciences as diverse as sociology and statistical mechanics all use the standard properties of probability, and the theorems asserted or claimed about probabilistic phenomena depend upon the formal properties that flow from the Kolmogorov axiomatization, plus possibly other special assumptions needed in particular applications. For example, the theory of stochastic processes is of increasingly great importance.

Because of the significance I shall attach to the relation of any interpretation of probability to the Kolmogorov axiomatization, it will probably be worthwhile to formulate the Kolmogorov approach. The axioms are based on three primitive concepts: a nonempty set X of possible outcomes, a family $\mathcal{F}$ of subsets of X representing possible events, and a real-valued function P on $\mathcal{F}$; for event A in $\mathcal{F}$; $P(A)$ is interpreted as the probability of A. It is important to note in this connection that events are always formally represented as subsets of the basic sample space X. (It is assumed throughout that X is nonempty.) The notion of an algebra of events is caught in the following definition.

DEFINITION 1. *$\mathcal{F}$ is an* algebra of events *on X if and only if $\mathcal{F}$ is a nonempty family of subsets of X and for every A and B in $\mathcal{F}$:*

1. $\sim A \in \mathcal{F}$;

2. $A \cup B \in \mathcal{F}$.

Moreover, if $\mathcal{F}$ is closed under countable unions, that is, if for

$$A_1, A_2, \ldots, A_n \ldots \in \mathcal{F} \cup_{i=1}^{\infty} A_i \in \mathcal{F},$$

then $\mathcal{F}$ is a σ-algebra on X.

Assuming the set-theoretical structure of X, $\mathcal{F}$, and P already described, we may now turn to the definition of probability spaces.

DEFINITION 2. *A structure $\mathcal{X} = \langle X, \mathcal{F}, P\rangle$* is a finitely additive probability space *if and only if for every A and B in $\mathcal{F}$:*

P1. *$\mathcal{F}$ is an algebra of events on X;*
P2. $P(A) \geq 0$;
P3. $P(X) = 1$;
P4. *If $A \cap B = 0$, then $P(A \cup B) = P(A) + P(B)$.*

Moreover, $\mathcal{X}$ is a probability space (without restriction to finite additivity) *if the following two axioms are also satisfied:*

P5. *$\mathcal{F}$ is a σ-algebra of events on X;*

P6. *If $A_1, A_2, \ldots$, is a sequence of pairwise incompatible events in $\mathcal{F}$, i.e., $A_i \cap A_j = 0$ for $i \neq j$, then*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Some more special probabilistic notions will be needed in later parts of this paper, but I shall defer until then their formal definition. Also, although I have included the countable cases for the sake of completeness in Definitions 1 and 2, I shall mainly ignore them in the sequel. These concepts are of central importance in advanced applications, but explicit consideration of them here does not add much of conceptual interest, and the notation and theorems will remain simpler if they are ignored. We can, if we desire, think of the sequel as being restricted to finite probability spaces, that is, to spaces in which the basic set X is finite. But, I would emphasize that the conceptual remarks are not at all bound by this imposition of finiteness.

On the assumption that any interpretation of probability must come to terms with the Kolmogorov set-theoretical approach as embodied in Definitions 1 and 2, we can ask ourselves how is that "coming to terms" to be expressed. There is a classical mathematical way of formulating the matter. We must be able to prove that the set-theoretical entities defined under the particular interpretation of probability are themselves either objects that satisfy Definition 2 or lead in a completely explicit way to the construction of objects that satisfy Definition 2. This rather abstract formulation of the representation problem is made more concrete by consideration of the classical interpretations of probability.

The place to begin is with Laplacean or classical probability. Laplace begins this way. "The first of these principles is the definition itself of probability, which, as has been seen, is the ratio of the number of favorable cases to that of all the cases possible." There are severe difficulties with the application of this definition; but it is clear how it leads to a representation in terms of probability spaces. We may incorporate the idea of Laplace's first principle in the following formal definition. (In the statement of this definition I use $K(A)$ for the cardinality of the set A.)

DEFINITION 3. *A structure $\mathcal{X} = \langle \mathcal{X}, \mathcal{F}, \mathcal{P} \rangle$ is a* finite Laplacean probability space *if and only if:*

P1. *X is a finite set;*

P2. *$\mathcal{F}$ is an algebra of events on X;*
P3. *For A in $\mathcal{F}$,*

$$P(A) = \frac{K(A)}{K(X)}.$$

It is apparent that the following theorem is a trivial consequence of this definition, but the theorem does express the way in which the Laplacean definition provides a strict interpretation of the set-theoretical formalization of probability.

THEOREM 1. *Any finite Laplacean probability space $\mathcal{X} = \langle X, \mathcal{F}, P\rangle$ is a finitely additive probability space in the sense of Definition 2.*

Almost everybody recognizes criticisms that can be leveled at the classical Laplacean definition. It is only when very strong principles of symmetry are satisfied that the definition in a strict sense can be applied. Most of Laplace's own work in probability did not use the principle. I mean by this that the detailed applications in astronomy and other complicated phenomena did not proceed from a strict application of the classical definition.

However, I find myself unable to agree with some of Popper's criticisms of the classical definition. These criticisms are expressed in *PI* (p. 36).

> ...mere possibilities could never give rise to any prediction. It is possible, for example, that an earthquake will destroy tomorrow all the houses between the 13th parallels north and south (and no other houses). Nobody can calculate this possibility, but most people would estimate it as exceedingly small; and while the sheer possibility as such does not give rise to any prediction, the estimate that it is exceedingly small may be made the basis of the prediction that the event described will not take place ("in all probability").
>
> Thus the estimate of the *measure* of a possibility—that is, the estimate of the probability attached to it—has always a predictive function, while we should hardly predict an event upon being told no more than that this event is possible. In other words, we do not assume that a possibility as such has any tendency to realise itself; but we do interpret probability measures, or "weights" attributed to the possibility, as measuring its disposition, or tendency, or propensity to realise itself...

It seems to me that in this discussion there is an abuse of the Laplacean notion of possibility. It is precisely the point that it is simply the enumeration of possibilities that are used in the definition of probability. I think

Popper can rightly object, as he does, that in many instances this is not what we do, or we do not know how to enumerate the possibilities. On the other hand, in the examples given by Laplace as paradigm cases of his definition, namely, games of chance, we do agree on the enumeration of possibilities, and we can claim that it is the enumeration of possibilities that provides the basis for the definition of probability. It is not a case, as Popper puts it, of assuming "that a possibility as such has any tendency to realise itself." It is just that the computation of probabilities is based upon the enumeration of possibilities. As we look deeper into the matter, we are willing to accept these enumerations in cases where certain implicit principles of symmetry are satisfied and not in other cases. My quarrel with Popper here is not a large one. I do think that the notion of possible outcomes or possible cases is a fundamental aspect of thinking about probability. It is also a fundamental aspect to tend to impose a uniform distribution on the set of possible outcomes. We do not do this because of any ideas about propensity itself or probability weights, but because we see no reason to treat one possible outcome in a manner different from another. I want to be perfectly clear on this point. I am not trying to defend the classical definition as a workable interpretation of probability. I am taking exception to Popper's remarks about the way in which possibility does enter in the classical definition. The important point for our discussion here is that a well-defined formal relation between the classical interpretation of probability and the set-theoretical approach is embodied in Definition 2. This formal relation is caught by Theorem 1.

Let us now take a quick look at the relative-frequency interpretation of probability. I shall first state several formal notions and a theorem, but then indicate how the models brought under the framework of probability spaces by the theorem are not fully satisfactory. However, I shall not enter into the formal details of how the definitions are to be more restricted. These are standard matters in the discussion of relative-frequency theory. The point is to show how one is led from the relative-frequency interpretation in a formal way to models of Definition 2.

An (*infinite*) *sequence* is a function whose domain of definition is the set of positive integers. If s is a sequence then $s(n)$ (or in a common notation: s_n) is the n^{th} *term* of the sequence s. A sequence of *positive integers* is a sequence all of whose terms are positive integers.

Let s be a sequence of real numbers. Then the $\lim_{n\to\infty} s_n = k$ if and only if for every $\in > 0$, there is an integer N such that for all integers n if $n > N$ then

$$|s_n - k| < \in .$$

DEFINITION 4. *Let s be a sequence and let $\mathcal{F}$ be the family of all subsets of $R(s)$, the range of s. Let t be the function defined on $\mathcal{F}\times\omega$ where ω is the set of all positive integers, such that for all A in $\mathcal{F}$,*

$$t(A,n) = K\{i : i \leq n \ \&\ s_i \in A\}.$$

The number $t(A,n)/n$ is the relative frequency of A in the first n terms of s. *If the limit of the function $t(A,n)/n$ exists, then this limit is* the limiting relative frequency of A in s.

THEOREM 2. *Let s be a sequence and let $\mathcal{F}$ an algebra of events on $R(s)$ such that if $A \in \mathcal{F}$ then the limiting relative frequency of A in s exists. Let P be the function defined on $\mathcal{F}$ such that for every A in $\mathcal{F}$*

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

Then $\langle R(s),\ \mathcal{F},P\rangle$ is a finitely additive probability space.

The proof of Theorem 2 requires more argument than the proof of Theorem 1, but it is straightforward in terms of standard facts about the limits of sequences and will be omitted here.

I emphasize that to have a realistic relative-frequency theory, the conditions of Theorem 2 need to be strengthened. Many sequences satisfy the hypothesis of Theorem 2 and thus generate a finitely additive probability space, but we would not at all be willing to accept them as falling within the framework of what we intuitively consider to be probabilistic phenomena. For example, the deterministic sequence consisting of alternating 1's and 0's would satisfy Theorem 2 and the event of a 1 occurring would be 1/2, and the event of a 0 occurring, 1/2; but, clearly, no reasonable notion of probability in an intuitive sense would admit such a sequence. The point of the present discussion, however, is not disturbed by this aspect of things. I am interested only in how we formulate a formal relation between a relative-frequency theory and the notion of probability space embodied in Definition 2.

I turn now to a brief exposition of the subjective theory of probability and the way in which it formally provides an interpretation of Definition 2. I shall restrict my analysis of the subjective theory to a simple example to avoid technical complexities. The spirit of the axioms embodied in the definition given below places restraints on qualitative judgments of probability, which on the one hand seem intuitively sensible and on the other hand seem sufficient to guarantee the existence of a numerical probability measure in the sense of Definition 2. The subjective aspect enters directly in the sense that the qualitative relation is meant to reflect the qualitative judgments of subjective probability: $A \succeq B$ if and only if A is judged subjectively at least as probable as B.

DEFINITION 5. *A structure* $\mathcal{X} = \langle X, \mathcal{F}, \succeq\rangle$ *is a* finite qualitative probability structure with equivalent atoms *if and only if* X *is a finite set,* $\mathcal{F}$ *is an algebra of events on* X, $\succeq$ *is a binary relation on* $\mathcal{F}$, *and the following axioms are satisfied for every* A, B, *and* C *in* $\mathcal{F}$:

1. *The relation* $\succeq$ *is a weak ordering of* $\mathcal{F}$;
2. *If* $A \cap C = \emptyset$ *and* $B \cap C = \emptyset$, *then* $A \succeq B$ *if and only if* $A \cup C \succeq B \cup C$;
3. $A \succeq \emptyset$;
4. *Not* $\emptyset \succeq X$;
5. *If* $A \succeq B$ *then there is a* C *in* $\mathcal{F}$ *such that* $A \succeq B \cup C$ *and* $B \cup C \succeq A$.

The first four axioms are standard axioms that originate with de Finetti (the symbol $\emptyset$ stands for the empty set); the fifth axiom is the structural axiom that implies the equivalence of the atoms; the exact theorem that can be proved is the following:

THEOREM 3. *Let* $\mathcal{X} = \langle X, \mathcal{F}, \succeq\rangle$ *be a finite qualitative subjective probability structure with equivalent atoms. Then there exists a probability measure* P *in the sense of Definition 2 such that for every* A *and* B *in* $\mathcal{F}$

$$P(A) \geq P(B) \textit{ if and only if } A \succeq B.$$

Moreover, there are at most two equivalence classes of atomic events in $\mathcal{F}$; *and if there are two rather than one, one of these contains the empty event.*

The proof of this theorem I shall omit.[2]

It is not my purpose in this paper to defend any one of the three views of probability I have sketched above. Rather, in the present context I want to distinguish the three classical views sketched above from the propensity interpretation advocated by Popper on the grounds that I do not see what the corresponding theorem for the propensity interpretation is. I have gone to some length to make this point, because I think it is an important one about the propensity interpretation. I very much agree with Popper that there is much that is attractive in the idea of probability as propensity. What I find difficult to envisage, and what I find missing in his own discussions of the propensity interpretation, is the more explicit formal characterization of the propensity interpretation that permits us to prove a theorem like Theorem 1, 2 or 3. Until an interpretation of

[2] The elementary proof is to be found in Suppes (1969a), pp. 7–8.

probability is given sufficient systematic definiteness to permit the proof of such a theorem, it seems fair to say that it is still at a presystematic level, and no clear concept has as yet been explicated.

Popper tells us in *PI* that he was especially led to the propensity concept by the problem of interpreting the use of probability in quantum theory. He felt that the Bohr-Heisenberg interpretation was inextricably bound up with the subjectivistic interpretation of probability. On the other hand, the difficulty of the relative-frequency theory lay in providing an appropriate straightforward interpretation of the probability of singular events. A number of his remarks in this connection seem to me sensible, as for example, his insistence that the so-called "problem of the reduction of the wave packet" is a problem inherent in every probabilistic theory, and not special to any particular interpretation.

There are a number of tantalizing remarks about the propensity view in *PI*. In several places Popper compares propensities to forces in Newtonian physics. As he puts it, "there is an analogy between the idea of propensities and that of forces." However, I would again raise the same question I have been raising. Already in the case of Newtonian forces there are explicit formal laws that these forces are required to obey: the laws of addition of forces and also the more special laws for internal forces in a system of particle mechanics; namely, the law that the force exerted by one particle on another be equal and opposite, and also the law that the direction of these two internal forces be along the line connecting the position of the two particles. I find no systematic laws whatsoever that the propensity interpretation is to satisfy, except the formal laws of probability already embodied in Definition 2.

In other passages, Popper indicates the close relation between the propensity interpretation and the relative-frequency interpretation, but again I would want to press the point and ask if there is indeed a formal difference between the two and, if so, what it is.

Of the three views I have sketched above, the relative-frequency and subjective views each provide a sharply defined formal theory that does lead to an interpretation in the formal sense of the axioms of Definition 2. The classical theory also provides such an interpretation, but it is weaker and less interesting.

Toward the close of *PI*, Popper says the following:

> ...what I propose is *a new physical hypothesis* (or perhaps a metaphysical hypothesis) analogous to the hypothesis of Newtonian forces. It is the hypothesis that every experimental arrangement (and therefore every state of a system) generates physical propensities which can be tested by frequencies. This

> hypothesis is testable, and it is corroborated by certain quantum experiments. The two-slit experiment, for example, may be said to be something like a crucial experiment between the purely statistical and the propensity interpretation of probability, and to decide the issue against the purely statistical interpretation.

What troubles me about this passage is the vagueness of his new physical hypothesis in contrast to the sharpness of formulation of the hypothesis of Newtonian forces. From what he says it is also not clear how the two-slit experiment provides a crucial experiment between the relative-frequency and propensity interpretations. Indeed, I have found it difficult to try to infer what formal properties the propensity interpretation is supposed to have from consideration of the two-slit experiment.

Let me sum up the situation as I see it in three points.

1. Much of what Popper says about the use of probability in quantum mechanics and the way he has used the idea of propensity to say these things seem eminently sensible to me.

2. I find the systematic case for the propensity interpretation badly worked out and not at all in a state comparable to that of the classical, relative-frequency or subjective interpretations.

3. I recognize at the same time that the subjective theory, especially in its simply providing a qualitative ordering relation, has not provided an interpretation at a very deep level. I do not wish to defend the adequacy of the subjective interpretation in any fundamental way. It does stand in sharp contrast, however, to the propensity interpretation because there does exist a systematic body of analysis and resulting theorems that can be proved about the subjective view. Until such an analysis and resulting theorems are produced for the propensity interpretation, I find it impossible to embark upon a more thoroughgoing critique.

2. QUANTUM MECHANICS AS A STATISTICAL THEORY

Popper has written extensively on the conceptual nature of quantum mechanics. I shall not cite here the many references, for these are available in the general bibliography. I agree with much of what he has had to say about quantum mechanics as a statistical theory. He has had many sane and sensible things to say in his analyses and criticisms of the doctrines advanced by physicists. Our points of agreement are too many to enumerate, although I cannot help mentioning my pleasure in his recent article on Birkhoff and von Neumann's interpretation of quantum mechanics (Popper, 1968). He points out the conceptual inadequacy of the argument given in Birkhoff and von Neumann's famous article in a way that is perhaps the clearest of anything I have seen in print. I shall not review the details of the argument here, but note that he shows how unsatisfactory is their claim that quantum mechanics uses a nondistributive lattice. It is not the result that is so unsatisfactory, but the total lack of serious argument for their position.

I could list other points of agreement, but the more constructive and useful thing is to focus on the major issues where I find myself in disagreement with Popper, or where I do not think he has pushed hard enough or dealt as yet sufficiently explicitly with matters of central importance.

The central theme of what I want to say can be posed as a question. Is indeed quantum mechanics a genuinely statistical theory? By this, I do not question whether there are many statistical aspects of quantum mechanics, but rather, can quantum mechanics as a theory be regarded as a statistical theory in the way that classical statistical mechanics, population genetics or theories of mental testing are statistical theories? It seems to me that much in Popper's writings indicates that he would want to make this claim. I shall not try to document the many places where he discusses these matters, but I would refer the reader especially to his recent article, "Quantum Mechanics Without 'The Observer'."[3] In this article Popper sets forth 13 theses about quantum mechanics. It is not possible to examine each of these theses and to comment on them, especially as to how each thesis relates to the view of quantum mechanics as a statistical theory. I shall begin by concentrating on the interpretation of the Heisenberg uncertainty relations, and then go on from there to problems about probability that are not explicitly treated by Popper.

To begin with, if one starts from the idea that quantum mechanics is a statistical theory, as Popper evidently does, a first point of peculiarity about the Heisenberg principle needs to be mentioned. The principle asserts that the product of the standard deviations of two noncommuting variables is always greater than some positive constant. In the particular example of position and momentum it is asserted that the products of the standard deviations of the position and momentum of a particle at a given time are always greater than a certain fixed constant, which is positive. A statement of this kind can be derived in many classical theories. In fact, it will be true in any classical theory in which we are dealing with a nondegenerate joint distribution of at least two random variables. Nearly the first thing we would want to do is ask about a closer relation between the two variables. We would want information about the covariation of the two random variables and their possible causal relation—that causal

[3] K. R. Popper (1967). Hereinafter cited as QM.

relation either being between the variables or due to a common cause. The standard statistical way in which these matters are studied would be in terms of looking at the covariance or the correlation of the two variables. (Because the notion of correlation is familiar in a wide range of scientific disciplines, let us deal with the correlation and recall that the correlation of two random variables is defined as the covariance divided by the product of the standard deviations.) Given that the product of the standard deviations is greater than some fixed constant, we can still produce examples in which the correlation has the entire range from -1 to 1; in particular, examples for which the correlation between the values of the random variables is 1, and also cases for which the correlation of the random variables is 0. From a general statistical viewpoint, it is often more important and almost always at least as important to know whether the random variables are independent or highly correlated, as it is to know that the product of their standard deviations is greater than some constant. When Popper talks about quantum mechanics as a statistical theory, he is talking, it seems to me, with that surprise evinced by those who look at quantum mechanics from the standpoint of classical physics—surprise that the theory brings within its purview certain statistical relations and denies at the theoretical level the determinism so characteristic of classical physics. Looked at from the standpoint of standard statistical theories, the surprise about quantum mechanics is rather different. The first glance would be something like the one I have sketched. The surprise is that natural questions are not asked or discussed. Popper's own neglect of these standard questions of covariation or correlation is a reflection that he has not really taken seriously as yet the rethinking of quantum mechanics as a statistical theory. What Heisenberg, for example, has had to say about these matters would make the hair of any right-thinking statistician stand on end.

I emphasize the importance of these questions of correlation. If, for example, the Heisenberg uncertainty relation is satisfied by position and momentum in a given direction at a given time, we would be enormously surprised if the correlation between position and momentum were one. This would indicate a deterministic relation between the two that would be most disturbing to most physicists. I stress, however, that such a model is mathematically consistent with the Heisenberg relations. This is an obvious and elementary fact of statistics. It is of course not my claim that such an interpretation is consistent with the actual empirical data of quantum mechanics or with the theory taken in a larger framework than that of the simple statement of the Heisenberg principle.

That physicists and Popper as well do not really take seriously their claim that quantum mechanics is a statistical theory is evident from the complete absence of discussion of the problems of correlation just raised. In Popper's case, I suspect that he has simply been caught up in the discussions of physicists and has tried to respond in a direct way to the many kinds of things they have had to say; he has not looked at the problems from the standpoint of a genuine statistical theory.

Let me now turn to the second part of my remarks in this connection. There are good reasons why the questions I have raised are not raised. There are many ways of explaining what the reason for failure is. The essential idea, however, is that quantum mechanics is not a standard statistical theory—it is a peculiar, mystifying, and as yet, poorly understood radical departure from the standard methodology of probability and statistics. There is as yet no uniform agreement on how the probabilistic aspects or statistical aspects of quantum mechanics should be formulated. But it is widely agreed that there are unusual problems that must be dealt with and that do not arise in standard statistical theories of the sort I mentioned earlier. In fact, the kind of problems I now want to raise do not, so far as I know, exist in any other scientific theories of any scientific discipline.

The difficulty is that when the standard formalism of quantum mechanics is used, the joint distribution of noncommuting random variables turns out not to be a proper joint distribution in the classical sense of probability. These ideas have been discussed now by a good many people, and I shall not quote chapter and verse here. I am sure that Popper is familiar with several of these discussions, although I have been a little surprised not to find more explicit comment on these matters in his own writings. I do think the difficulties raised by the nonexistence of joint distributions within the framework of the standard formalism are the most direct challenge to a straightforward interpretation of quantum mechanics as a standard statistical theory.

To have a concrete instance in front of us, I give an example that I computed some years ago (Suppes, 1961), but I emphasize that these matters have been discussed by many people and general proofs of the impossibility of having proper joint distributions within the classical framework have been given by several people.

Consider the momentum and position random variables P and Q. The characteristic function $\varphi(t, u)$ is defined by:

$$(1) \qquad \varphi(u, v) = E(e^{iup+ivq}).$$

Using the Hilbert space formulation, let (ψ, ψ) be the inner product of a state with itself. Following the usual formalism, the expectation $E(R)$ of an operator R when the quantum mechanical system is in state ψ is simply $(\psi, R\psi)$. In view of (1) the characteristic function $\varphi(u,v)$ for the joint distribution of P and Q is given by:

$$\varphi(u,v) = (\psi, e^{i(up+vq)}\psi). \tag{2}$$

We then have from (1) and (2) by Fourier inversion:

$$f(p,q) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i(up+vq)}(\psi, e^{i(up+vq)}\psi)du\, dv. \tag{3}$$

For canonically conjugate operators P and Q, i.e., $PQ - QP = \hbar/i$, it may be shown that (2) simplifies to[4]

$$\varphi(u,v) = \int \psi^*(q - \frac{1}{2}\hbar u)e^{ivq}\psi(q + \frac{1}{2}\hbar u)dq$$

and so by Fourier inversion

$$f(p,q) = \frac{1}{2\pi}\int \psi^*(q - \frac{1}{2}\hbar u)e^{-iup}\psi(q + \frac{1}{2}\hbar u)du. \tag{4}$$

As is well known in probability theory not every characteristic function determines a proper probability distribution, and this is the difficulty with (4). (The expression given by (4) for the joint density was first proposed by Wigner (1932) and the derivation just sketched follows Moyal (1949).)

Let us now look at a simple example, the harmonic oscillator in the ground state and also in the first excited state.

Ground state. The potential energy is given by

$$V(x) = \frac{1}{2}Kx^2,$$

and the time-independent wave equation is

$$-\frac{h^2}{2m}\frac{d^2\psi(x)}{dx^2} + \frac{1}{2}Kx^2\psi(x) = E\psi(x).$$

The solution of this equation in terms of Hermite polynomials is familiar from the literature. In the lowest energy state H_0

$$\psi(x) = (\frac{\alpha}{\pi^{1/2}})^{\frac{1}{2}} \exp(-\frac{1}{2}\alpha^2x^2), \tag{5}$$

[4] Henceforth the range of integration is understood to be $(-\infty, \infty)$ and notation for it is omitted.

where

$$\alpha^2 = \sqrt{\frac{Km}{\hbar^2}}.$$

Thus

$$|\psi(x)|^2 = \frac{\alpha}{\pi^{1/2}} e^{-\alpha^2 x^2} \tag{6}$$

which is a normal density with mean zero and variance $\sigma^2 = 1/2\alpha^2 = \hbar/2\sqrt{Km}$.

We now apply (4) and (5) to obtain the joint distribution of momentum and position. For convenience of calculation, we replace p by the propagation vector $k = p/\hbar$. We have at once:

$$\begin{aligned} f(k,x) &= \tfrac{1}{2\pi} \int \psi^* \left(x - \tfrac{u}{2}\right) e^{-iku} \psi \left(x + \tfrac{u}{2}\right) du \\ &= \tfrac{1}{2\pi} \left(\tfrac{\alpha}{\pi^{1/2}}\right) \int \exp \left[\alpha^2 \left(x^2 + \left(\tfrac{u}{2}\right)^2\right)\right] e^{-iku} du \\ &= \tfrac{1}{2\pi} \left(\tfrac{\alpha}{\pi^{1/2}}\right) e^{-\alpha^2 x^2} \tfrac{\pi^{1/2}}{\alpha/2} \exp \left[-\tfrac{k^2}{4(\alpha/2)^2}\right] \\ &= \tfrac{1}{\pi} \exp \left(-\alpha^2 x^2 - \tfrac{k^2}{\alpha^2}\right). \end{aligned}$$

First excited state. We have from the literature

$$\psi(x) = \left(\frac{4\alpha^3}{\pi^{1/2}}\right)^{1/2} x \exp \left(-\frac{1}{2}\alpha^2 x^2\right),$$

whence

$$|\psi(x)|^2 = \frac{4\alpha^3}{\sqrt{\pi}} x^2 e^{-\alpha^2 x^2}.$$

Applying now (4) and (5), and again replacing p by the propagation vector $k = p/\hbar$, we have:

$$(7)\ f(k,x) = \left(\tfrac{1}{2\pi}\right)\left(\tfrac{4\alpha^3}{\sqrt{\pi}}\right) \int \left(x^2 - \left(\tfrac{u}{2}\right)^2\right) \exp\left[-\alpha^2 \left(x^2 + \left(\tfrac{u}{2}\right)^2\right)\right] e^{-iku} du.$$

Integrating (7) we obtain

$$f(k,x) = \tfrac{4}{\pi} \left[\exp \left(-\alpha^2 x^2 - \tfrac{k^2}{\alpha^2}\right)\right] \left(\alpha^2 x^2 + \tfrac{k^2}{\alpha^2} - \tfrac{1}{2}\right),$$

and the function $f(k,x)$ is negative for those values of k and x such that

$$\alpha^2 x^2 + \tfrac{k^2}{\alpha^2} < \tfrac{1}{2},$$

which means that $f(k,x)$ is not a proper joint density.

To my mind the problems posed by this elementary example and others like it, as well as general results about the impossibility of having a joint distribution, constitute the really central question of how to treat quantum mechanics as a statistical theory. This is not the proper place to examine possible viewpoints or to review some of the few proposals that have been made; for example, that by Margenau and his collaborators to adopt the special joint distribution that makes noncommuting random variables independent. What I consider important in the present context is to bring to the surface the deep-running nature of the difficulties of interpreting quantum mechanics as a standard statistical theory.

The thirteenth and last thesis of QM is "the peculiarity of quantum mechanics is the principle of the superposition of wave amplitudes—a kind of probabilistic dependence... that apparently has no parallel in classical probability theory." It is also part of this last thesis to say that both classical physics and quantum physics are indeterministic. What I would urge upon Popper is not the view of the peculiarity of quantum mechanics in terms of the principle of the superposition of wave amplitudes, but rather, the peculiarity of quantum mechanics as a nonstandard statistical theory. Given the wide applicability in all ordinary domains of science of the standard statistical theory and methodology, it is surprising and intellectually unsettling to encounter the fundamental difficulties that seem to be present in quantum mechanics. These difficulties disturb a much deeper level of scientific methodology than do any mere issues of determinism.

In my judgment, these formal difficulties of interpreting quantum mechanics as a standard statistical theory will turn out to be the most revolutionary aspect of the theory. My own historical sense is that these difficulties will come to play the same fundamental role in the foundations of physics and probability that the three classical problems of the Greeks have played in the foundations of mathematics. We now all accept that we cannot trisect an angle or find a square whose area is equivalent to that of a given circle by elementary means. I do not think we have as yet digested in any deep and serious way the profound ramifications of the nonstandard statistical character of quantum mechanics.

23

PROBABILISTIC CAUSALITY IN QUANTUM MECHANICS

I want to begin by expressing my pleasure at being able to contribute to a symposium in honor of Jack Good. We have known each other more years now than I care to remember. Over this long period I have learned much from his papers and from conversations about many different but related subjects. To make the point more precise for the present talk, about two decades ago when I became seriously interested in probabilistic causality and committed myself to developing a series of lectures on the subject, which were given in the summer of 1966 in Vaasa, Finland, among the few serious publications on the topic I found were Jack's important earlier papers (Good, 1961/1962). Recently I have commented on some of our points of disagreement about probabilistic causality (Suppes, 1988). On the other hand, our areas of agreement about probabilistic causality certainly exceed the filigree of differences.

What I want to do in this lecture is to explore what I regard as the absence of probabilistic causality in quantum mechanics.

1. OVERVIEW

The minor fact to be stressed is that if we avoid noncommuting variables in quantum mechanics, then probability is classical. In other words, any

*Reprinted from, *Journal of Statistical Planning and Inference*, **25** (1990), 293-302.

finite family of commuting variables, for example, the $3n$ position coordinates of n particles at a given time have a classical joint probability distribution.

A fact that has been much emphasized in the literature is that noncommuting variables in general do not have joint probability distributions, and thus a straightforward classical probabilistic theory of quantum phenomena is not possible. A typical example of two noncommuting observables that do not have a proper joint probability distribution is provided by the case of position and momentum for the one-dimensional harmonic oscillator in the first excited state (Suppes, 1961). It may be shown by standard methods that when we replace momentum p by the propagation vector $k = p/\hbar$ then the joint 'density' we obtain is:

$$(1)\quad f(k,x) = \left(\tfrac{1}{2\pi}\right)\left(\tfrac{4\alpha^3}{\sqrt{\pi}}\right) \int \left(x^2 - \left(\tfrac{u}{2}\right)^2\right) \exp\left[-\alpha^2 \left(x^2 + \left(\tfrac{u}{2}\right)^2\right)\right] e^{-iku} du.$$

Performing the integration of (1) we obtain

$$(2)\qquad f(k,x) = \tfrac{4}{\pi} \left[\exp\left(-\alpha^2 x^2 - \tfrac{k^2}{\alpha^2}\right)\right] \left(\alpha^2 x^2 + \tfrac{k^2}{\alpha^2} - \tfrac{1}{2}\right).$$

We note at once that the function $f(k,x)$ is negative for those values of k and x such that

$$\alpha^2 x^2 + \frac{k^2}{\alpha^2} < \frac{1}{2},$$

and thus we see that $f(k,x)$ is not a proper joint probability density. Such examples are easily multiplied, but our main concern here is of another sort.

The important fact, in many ways, is not the difficulty about noncommuting variables not having joint probability distributions but rather that the use of probability in quantum mechanics is very limited. The terminology for describing the situation is not standard. My own preference is to say that quantum mechanics provides only a theory of the mean. Let me explain more precisely what I mean by this. There are, of course, many different marginal distributions that can be computed in quantum mechanics, but when we follow the Schrödinger equation and obtain in the time-dependent case a ψ function for each time t, then the square of this ψ function properly normalized gives us the distribution of the position variables, for example, at a given time t. This marginal distribution I call the mean distribution because it reflects the distribution of position at time t without any consideration of the sample paths or prior positions of the particles. Intuitively, we may think of the mean distribution as resulting from averaging over all possible sample paths up to time t. I shall restrict myself primarily to position variables just because the overlap with ordinary probability theory here is quite straightforward but also because, as is often remarked, all measurements in quantum mechanics can be reduced to measurements of position at a given time.

Thus from the standpoint of stochastic processes, not quantum mechanics, what we get in quantum mechanics is just the mean distributions at a given time. We get no autocorrelations or other features relating the positions of a given particle at different times. The absence of such autocorrelations or other information about the behavior of a particle at different times means that in the ordinary sense there is no probabilistic causality in quantum mechanics, for the essence of probabilistic causality is to relate behavior at one time to behavior at another time.

2. SOME EXAMPLES

The method for computing mean distributions is in principle straightforward, though difficult in particular examples. First, to get the expectation of an operator $\boldsymbol{A}$ in the standard approach

$$E(\mathbf{A}) = (\psi, A\psi)$$

where (ψ, ψ) is in the usual Hilbert space formulation the inner product of a state with itself.

To get the distribution of an observable $\boldsymbol{A}$ rather than just the expectation of $\boldsymbol{A}$, we replace $\boldsymbol{A}$ by e^{iuA} and obtain the characteristic function:

$$\varphi(u) = E(e^{iuA}) = (\psi, e^{iuA}\psi).$$

We then obtain the distribution of x in the form of the density $f(x)$ by taking the Fourier transform of φ.

Free particle. A classic simple example is that of the free particle acted upon by no forces. In the case of the free particle, by using the time-dependent Schrödinger equation we just obtain the following expression, which shows that for each t, $\boldsymbol{X}$ is a normally distributed random variable:

$$f(x,t) = \left(2\pi\left(\sigma_0^2 + \frac{\hbar^2 t^2}{4m^2\sigma_0^2}\right)\right)^{-1/2} \exp\left(-\frac{x^2}{2(\sigma_0^2 + \hbar^2 t^2/4m^2\sigma_0^2)}\right)$$

where

$$\sigma_0^2 = \mathrm{Var}(\boldsymbol{X}) \qquad \text{at } t = 0.$$

Notice that it is a feature of $f(x,t)$ that the variance increases in either direction of time.

From a stochastic standpoint it is natural to ask immediately additional questions about the process that would ordinarily be thought to be underlying the mean density $f(x,t)$. We would ordinarily ask such questions as: (1) Is the process Markovian, that is, for

$$\boldsymbol{X}_{t_1},\cdots,\boldsymbol{X}_{t_n} \text{ with } t_1 < t_2 < \cdots < t_n$$

is

$$F(\boldsymbol{X}_{t_n}|\boldsymbol{X}_{t_{n-1}},\cdots,\boldsymbol{X}_{t_1}) = F(\boldsymbol{X}_{t_n}|\boldsymbol{X}_{t_{n-1}})?$$

The answer is clearly that it is not determined by the axioms of quantum mechanics alone.

(2) Is the process Gaussian, i.e., does any finite sequence of position random variables have a multivariate normal distribution? Again the answer in no sense is determined by the standard axioms of quantum mechanics.

(3) Even when the answers are negative to questions (1) and (2), we may still ask, can we compute the autocorrelation function

$$\Gamma(t_1,t_2) = E(\boldsymbol{X}_{t_1},\boldsymbol{X}_{t_2})?$$

But again we have the same answer: no determination from the standard axioms of quantum mechanics.

We return to the point that only the mean density $f(x,t)$ is given by quantum mechanics, but it is also important not to think of quantum mechanics as just the mean distribution. There is important phase information given in the ψ function and we can use the ψ function also to compute the mean distribution $f(p,t)$ of the momentum. What we cannot do is get beyond these mean distributions.

Harmonic oscillator. We get the same kind of results for the one-dimensional linear harmonic oscillator already mentioned in the previous section. We obtain by standard methods the following mean density for position:

$$f(x,t) = \frac{\alpha}{\pi^{1/2}}e^{-\alpha^2(x-a\cos\omega t)^2}.$$

We can ask the same questions about the extension of this mean density. Is the process Markovian? Is it Gaussian? Can we compute the autocorrelation functions? And again we get the same answers. Nothing is determined directly by quantum mechanics.

3. STOCHASTIC EXTENSIONS OF QUANTUM MECHANICS

My central thesis, as is already perhaps evident, is that quantum mechanics is consistent with various stochastic extensions if we ignore computations on noncommuting variables.

Much of the view I am advocating is derived from the work of Edward Nelson (see particularly Nelson, 1967), but on one central point I do disagree with Nelson. His assumption of a detailed Brownian motion to provide a full dynamics for the free particle leads him to claim that the stochastic process view of the free particle is mathematically inconsistent with quantum mechanics, although there are no observable differences.

In my view his mistake is to compare the $\mathbf{X}_t$ of quantum mechanics with $x(t)$ of Brownian motion. Nelson says (1967, p. 39) that in quantum mechanics for the free particle

$$\boldsymbol{X}(\frac{t_1+t_2}{2}) = \frac{\boldsymbol{X}(t_1)+\boldsymbol{X}(t_2)}{2} \;,$$

but of course nothing so simple holds for $x(t)$ of Brownian motion. My rejoinder is that in quantum mechanics, adding $\boldsymbol{X}(t_1)$ and $\boldsymbol{X}(t_2)$ does not make any real sense, for we are not talking about the same particle, we are only dealing with the mean distribution with no particle identification across time possible. Thus, the equation given above is not a correct analysis of the free particle from a quantum mechanical standpoint, quite apart from any problems of measurement.

The view of quantum mechanics as giving only a theory of the mean is also a way of explaining why noncommuting variables are noncommuting. If position and momentum, for example, had a joint distribution, then we would, as in ordinary probability theory, anticipate we would be able to compute the covariance $\mathrm{Cov}(\boldsymbol{X}_t, \boldsymbol{P}_t)$. But in quantum mechanics we get instead the Heisenberg Uncertainty Principle:

$$\mathrm{Var}(\boldsymbol{X}_t)\mathrm{Var}(\boldsymbol{P}_t) \geq \text{constant} > 0,$$

and we are able to say nothing about the covariance, the natural quantity we would expect to study and the one we would expect to lead to a causal relation.

On the other hand, the quantum mechanical theory of the mean is at the right level. We are able to compute just about all that we can observe. A complete theory in the sense of Brownian motion does not seem to lead to any new and testable predictions.

On the other hand, because the theory of Brownian motion supplies a classical physical interpretation of quantum mechanics, it is surprising that the efforts have been as insubstantial as they have to extend this theory to quantum mechanical phenomena. Nelson's work almost stands alone. (In a moment I shall examine some deeper reasons why such a program in the framework of continuous-time Markov processes is not likely to be successful.)

I want to emphasize that the efforts of Nelson (1967, 1985) and others to extend quantum mechanics to stochastic mechanics by adding assumptions that extend quantum mechanics to being a Markovian diffusion process are philosophically important, because they provide a clear *dynamical* interpretation of quantum mechanics. As indicated in the earlier discussion, this extension has not necessarily always been looked upon by Nelson as an extension, but I think that that is the appropriate way to look at it from the conceptual standpoint of this paper. Earlier, Nelson (1966) derived the Schrödinger equation from Newtonian mechanics. More generally, what can be done is to characterize a classical diffusion process in the sense of stochastic mechanics that is compatible with the Schrödinger equation.

A good recent analysis of these matters is to be found in Yasue (1981). The line of argument in more detail goes as follows: Let $\psi(x,t)$ be a solution of the one-dimensional Schrödinger equation. This solution admits a polar decomposition

$$\psi(x,t) = \exp(R(x,t) + \mathrm{i}S(x,t)).$$

Then the Markov diffusion process generated by the infinitesimal generator

$$\mathcal{G} = b(x,t) \cdot \mathrm{grad} + \frac{\hbar}{2m}\mathrm{div\ grad}$$

satisfies the classical Euler equation for classical dynamics generalized to stochastic processes. In the above equation the drift b is given by

$$b = \frac{\hbar}{m}(\mathrm{grad}\ R + \mathrm{grad}\ S).$$

There is nothing unique about this associated Markov diffusion process. In particular, as we shall see later, there are arguments against the process even being Markovian. On the other hand, as a first approximation it seems like an excellent move, and, above all, it provides a detailed dynamical interpretation of the behavior of particles through time, in a way that quantum mechanics by itself does not. Also, from the standpoint of a central aspect of the present paper, this extension to a Markov diffusion process is a way of providing a full-blown causal framework for the motion of particles, even if the sample paths are unobservable. In contrast, quantum mechanics by itself does not provide a causal framework in any direct sense. The only weak sense of causality is that derived from the Schrödinger equation, namely, the law of change of the mean distribution with time.

4. LOCALITY

It seems to be the fate of the questions surrounding quantum mechanics that straightforward approaches run into trouble. From the standpoint of the development of stochastic processes in the past four decades, it seems completely natural to think of providing a physical interpretation of quantum mechanics in terms of Markov diffusion processes on the intuitive idea that particles are in continual Brownian motion. There are some physical problems with the interpretation of Brownian motion, for example, the standard mathematical result that the sample paths of particles are continuous but nowhere differentiable. But this kind of technical result, which might be treated as a kind of idealization for the purposes of simplifying the theory, does not present an insurmountable barrier—at least not without some new kind of experimental evidence that shows these ideas are in error. This would mean showing that predictions of Brownian motion clearly violated experimental data.

But this is just what has happened. The work deals with causal questions concerned with locality. From a broad philosophical standpoint, the arguments concern the traditional problem of action at a distance, but the special twists and turns of locality in quantum mechanics are new.

The classical striking results in this arena are due to Bell (1964). The idea is to test the existence of a causal structure—what are called in quantum mechanics hidden variables—along the following lines. If there is a causal structure to quantum mechanics, that is, an appropriate causal hidden variable λ—the cause is called hidden because it is not in any direct sense observable, then there are classical results about λ from a causal standpoint that we would expect to hold. Figure 1 shows in the usual one-dimensional diagram of special relativity the space-time region from which λ would have to be operating. It is the intersection of the back light cones of particles A and B—I am thinking here of course of λ being a cause influencing the behavior of particles A and B. We can also think of a Bell-type experiment here in which we are measuring spin for particle A and for particle B. More generally we would think of A and B being the location of measuring equipment and we observe individual particles or a flux of particles at each of the sites. We would still think

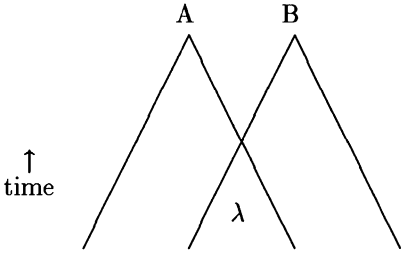

Figure 1. Physically possible location of hidden variable $\boldsymbol{\lambda}$.

here of individual particles because the analysis is conceptually simpler, even though some of the experiments would produce in fact collections of particles. We think of the measuring apparatus being such that along the axis connecting A and B we have axial symmetry and therefore we can describe the position of the measuring apparatus just by the angle of the apparatus A in the plane perpendicular to the axis. We shall use the notation w_A and w_B for these angles. I shall not attempt here a full technical analysis but only enough to give a sense of Bell's results and an informal description of their dire consequences for Markov diffusion processes. The basic form of the locality assumption is shown in terms of the following expectation:

$$(1) \qquad E(\boldsymbol{M}_A|w_A, w_B, \boldsymbol{\lambda}) = E(\boldsymbol{M}_A|w_A, \boldsymbol{\lambda}).$$

What this means is the expectation of the measurement $\boldsymbol{M}_A$ of spin of a particle in the apparatus in position A, given the two angles of measurement for apparatus A and B as well as $\boldsymbol{\lambda}$, is equal to the expectation without any knowledge of the apparatus angle w_B of B. This is a reasonable causal assumption and is a way of saying, looked at from the standpoint of special relativity, that what happens at B should have no direct causal influence on what happens at A because B is not in the back light cone of A. It violates strongly not only action at a distance in classical terms but even more in terms of special relativity. On the other hand, we have the following theoretical result for spin well confirmed in principle, for the case of when the measuring apparatuses are both set at the same angle:

$$(2) \qquad P(\boldsymbol{M}_A = -1 \mid w_A = w_B = \alpha \;\&\; \boldsymbol{M}_B = 1) = 1.$$

Note what is going on here. If the angles of the apparatus are set the same we have a deterministic result in the sense that the observation of spin at B will be the opposite at A, and conversely. Here we are letting 1 correspond to spin $\frac{1}{2}$ and -1 correspond to spin $-\frac{1}{2}$. There is a natural tension immediately observable between equations (1) and (2). The problem then is how to get a more specific test of whether or not locality is violated in quantum mechanics.

What Bell showed is that on the assumption there exists a hidden variable, four related inequalities can be derived for settings A and A' and B and B' for the measuring apparatus. I have reduced the notation here in the following way in writing the inequalities. First, instead of writing w_A I write simply A, and second, instead of writing $Cov(\boldsymbol{M}_A, \boldsymbol{M}_B)$ for the covariance, which in this case will be the same as the correlation, of the measurement at A and the measurement at B, I write simply $\boldsymbol{AB}$. With this understanding about the conventions of the notation we then have as a consequence of the assumption of a hidden variable the following set of inequalities, which in the exact form given here are due to Clauser, Horne, Shimony and Holt (1969):

$$\begin{aligned}
-2 \leq{}& \boldsymbol{AB} + \boldsymbol{AB'} + \boldsymbol{A'B} - \boldsymbol{A'B'} \leq 2,\\
-2 \leq{}& \boldsymbol{AB} + \boldsymbol{AB'} - \boldsymbol{A'B} + \boldsymbol{A'B'} \leq 2,\\
-2 \leq{}& \boldsymbol{AB} - \boldsymbol{AB'} + \boldsymbol{A'B} + \boldsymbol{A'B'} \leq 2,\\
-2 \leq{}& -\boldsymbol{AB} + \boldsymbol{AB'} + \boldsymbol{A'B} + \boldsymbol{A'B'} \leq 2.
\end{aligned}$$

What Bell showed is that quantum mechanics does not satisfy these inequalities, so they thus provide a clear test of the existence of local hidden variables, that is, causes that act in the appropriate local fashion. There have been a number of experimental tests in the past two decades, and it is fair to say that all of the experiments that have been accepted as valid have supported quantum mechanics.

Thus the theoretical and experimental results together present another and different body of evidence, very precisely focused against there being a standard causal structure which we may use to obtain a classical causal theory in the sense of hidden variables.

From the standpoint, though, of probabilistic causality in quantum mechanics, it is worth while to pursue a bit more these negative results about hidden variables. Fine (1982) proved that the Bell inequalities hold if and only if there exists a joint distribution of the random variables $\boldsymbol{A}, \boldsymbol{A'}, \boldsymbol{B}, \boldsymbol{B'}$. (As expected, the Bell inequalities are not sufficient for N>4.) This means that there is not a joint distribution in quantum mechanics because the inequalities are not satisfied. This is a familiar quantum mechanical story. In the present case the random variables $\boldsymbol{A}$ and $\boldsymbol{A'}$ are noncommuting, as are also the random variables $\boldsymbol{B}$ and $\boldsymbol{B'}$. From a conceptual standpoint, though, the noncommuting character of these pairs of variables is not in some sense quite the right result for, as we ordinarily think of the experiments, the random variables are random variables that are observed at different times. There is more to be said on this point but I am not going to do it here. Suppes and Zanotti (1981) proved that for the kind of random variables we are discussing here, namely, ones with values ±1, there exists a joint distribution of such random variables if and only if there exists a hidden variable such that

$$E(\boldsymbol{X}_1 \cdots \boldsymbol{X}_n|\boldsymbol{\lambda}) = E(\boldsymbol{X}_1|\boldsymbol{\lambda}) \cdots E(\boldsymbol{X}_n|\boldsymbol{\lambda}).$$

So the absence of joint distributions in quantum mechanics, a familiar result in the Bell kind of structure, implies according to this theorem that there can be no hidden variable $\boldsymbol{\lambda}$ and therefore, in the sense of local hidden variables, no causal theory.

This seems to contradict the earlier statement about extending quantum mechanics to Markov diffusion processes. Nelson (1985) shows in fact that the Markov stochastic mechanics does violate locality in the sense of Bell.

But all the possibilities are not lost with this result. As Nelson says, in principle a non-Markovian stochastic mechanics can be constructed that is consistent with locality. Such a construction has not been carried through in detail yet and certainly presents formidable difficulties, but there is certainly nothing in principle against such a construction. If such a construction can be carried through, probabilistic causality is restored for quantum phenomena, again by an appropriate extension of the mean theory of classical quantum mechanics. In principle there is also the possibility of introducing new concepts with respect to which the enlarged diffusion process is Markov, but this move takes us into uncharted waters searching for a fundamental conceptual extension of quantum mechanics distinct from hidden variable approaches.

I just want to conclude with one philosophical remark about non-Markovian processes. It is of course true that we are very used to thinking about the world in Markovian fashion. We find it hard to imagine that there is another kind of action at a distance operating, namely, action at a distance through time. Yet we all recognize that for many kinds of limited theories such action at a distance in time is absolutely essential. No one would propose today or consider it feasible in any sense whatsoever to be able to look at the structure of the brain of a person and give an account of the impact of past experience on that individual. Knowledge of actual events of the past is essential for a detailed understanding of the behavior of individuals. This does not just apply to the psychology of individuals—it is true for all kinds of other phenomena. Certainly we can improve our understanding of many complex phenomena, from the stock market to the weather, by taking account not just of the instantaneous state but of the past history as well. It is only when we think we have a final and fundamental theory of phenomena that we might be persuaded to say that the process must be Markov. Once we do not believe we have the fundamental theory then it is very natural to look for a theory that is non-Markovian in character. The concepts that we have formulated explicitly in a given theory we believe are not rich enough to catch all of the structure actually present in the entities being studied. We may indeed believe that it is not feasible in any foreseeable future to understand that structure completely but that we can make headway by looking at the history of the entity. Certainly we have that attitude in large-scale cosmology now. It seems to me that it is not philosophically surprising or upsetting to have to recognize that our best hope of providing a causal account of quantum phenomena may very well have to be non-Markovian in character.

PART V

PSYCHOLOGY

24

FROM BEHAVIORISM TO NEOBEHAVIORISM

1. DEFINITION OF NEOBEHAVIORISM

Nelson's (1975) detailed article on behaviorism and stimulus-response theory provides an opportunity for me to reformulate my own viewpoint and to clarify certain confusions in the informal discussion I gave of my formal results on the stimulus-response theory of finite automata (Suppes, 1969b).

The classical popular view of behaviorism is that it is strictly an operational theory, stated wholly in terms of observables. However, this is already not strictly true of detailed formal statements of the theory that go back as far as the late fifties (Estes and Suppes, 1959a; Suppes and Atkinson, 1960). In these mathematical formulations of behaviorism it was already apparent that the notion of stimulus was not directly observable and neither was the concept of conditioning. For example, in the experimental situations to which the theory was applied by a large number of investigators in the heyday of mathematical models of learning some fifteen years ago, it was even a standard trick to estimate the number of stimuli from the observed data but not to pretend to be able to identify the stimuli. If the stimuli were identified and if there was more than one stimulus in a presentation, then it was not possible in general

*Reprinted from *Theory and Decision*, 6 (1975), 269-285.

to identify precisely what stimuli were conditioned to what responses, and consequently to treat as an observable the relation of conditioning between stimuli and responses.

To clarify this situation in somewhat more detail, it is worth noting that the theories that were stated in terms of observables—for example, the stochastic models of Bush and Mosteller (1955) and the linear models of Estes and Suppes (1959a)—did not involve explicitly the concept of stimulus but only the two concepts of response and reinforcement. Even here, it was possible to insist on treating the concept of reinforcement as not being identifiable and there was some discussion of this matter in the work that Estes and I did in the late fifties. However, the canonical form of a stochastic model of learning that did not involve the concept of stimulus and that was applied to a wide variety of experiments was to treat the responses and reinforcements as observable, and to record the occurrence of each on a given trial. It is possible to insist that this model was not entirely observable because the probability of a response as opposed to the actual response was the theoretical entity of greatest importance, and this probability itself was not observable but could only be estimated approximately from experimental data.

Still, there is a point to the criticisms of the form of stimulus-response theory that was fashionable in the period I have just been discussing.[1] What came to be felt as the appropriate criticism within psychology of the work of those days was the absence of sufficient internal structure, the absence of anything of the complexity we intuitively associate with human mental abilities, especially the complex and subtle processes of memory and of language comprehension and production.

We are now in the era of neobehaviorism, which I would define in informal terms along the following lines. A theory of psychological phenomena is neobehavioristic if it recognizes as the essential observable data only stimulus conditions and responses, with both stimulus conditions and responses described in terms that are recognized as properly psychological. It is apparent that this is not meant to be a serious definition and I do not want here to attempt a serious definition. I am not even sure it is a worthwhile enterprise to do so. What I have in mind is excluding physiological or neurological observations and data, or biological data as, for example, gene structure. It is not that I think that psychology can be in

[1] In analyzing the observability of the various concepts or relations, I am not trying to split any hairs in a philosophical sense or to challenge the possibility of distinguishing between theoretical and observational terms. I am intending to give an analysis that is rather straightforward and not meant to be controversial in character.

any scientific sense properly and wholly autonomous from physiology or biology but rather that the most important psychological theories are to a large degree independent of physiology and biology. This independence is a thesis to which I shall return at the end of this article. For the present, I want to make the essential behavioral feature of neobehaviorism the retention of stimuli and responses as central on the one hand, and the introduction of unobservable internal structure as the 'neo' component on the other. Thus, in neobehaviorism as opposed to classical behaviorism it is quite appropriate to postulate a full range of internal structures, ranging from memory hierarchies to language production and language comprehension devices that cannot be, from the standpoint of the theory, directly observed.

2. THE REAL ISSUE: REINFORCEMENT

Nelson (1975) gives a detailed discussion of finite acceptors and finite transducers, and properly insists that the theory I set forth in Suppes (1969b) was a theory of finite acceptors. His central point is that my handling of responses in the earlier article does not permit an appropriate concept of internal state. He also points out that in my own informal discussion I moved too quickly to an identification of internal states and outputs. In his article he also emphasizes the possibility of having internal responses, and this is not really an issue between us. In fact, it seems to me there is no real issue between us. I agree with what he has to say about these matters and my intention is to focus on some of the central issues that he does not cover.

Surprisingly enough, perhaps the central omission from the standpoint of what I want to say in the present context is his absence of discussion of reinforcement. In my 1969 article I used a concept of *determinate* reinforcement. A reinforcement is determinate when the correct response is indicated after the actual response has occurred. For example, if I ask a child the sum of $7 + 5$, then a determinate reinforcement would be giving the correct answer, 12, when he gave an incorrect answer. An example of nondeterminate reinforcement would be simply to tell him that the answer was incorrect and to ask him to try again. When determinate reinforcement is used, it is clear that in some sense the responses have to be observable in order to correct each incorrect response. The theorem that I stated about finite automata in my 1969 paper used an assumption of determinate reinforcement. It seems to me that it is this assumption of determinate reinforcement rather than any of the informal remarks I made about responses being observable or internal states being identifiable with outputs that is really the central feature and the central limitation of the main theorem of that paper.

From the standpoint of giving an account of complex learning, especially in natural settings as opposed to simple laboratory situations, it was clear to me before the 1969 paper was published that the most essential extension was to obtain similar results with nondeterminate reinforcement. This problem was tackled in conjunction with my former student, William Rottmayer, and detailed results are embodied in his 1970 dissertation. An informal and rather brief statement of the results appears in a survey article we published on automata (Suppes and Rottmayer, 1974).

Because the formal statement of stimulus-response theory with nondeterminate reinforcement is rather complicated, I shall give only a brief informal statement similar to that in Suppes and Rottmayer (1974). Before doing so, let me formulate a canonical class of problems that can be used for intuitive reference in digesting the content of the individual axioms. The organism is presented with a potentially infinite class of stimulus displays, for example, line drawings. A subclass of the entire class is characterizable by a finite automaton. The problem for the learner is to learn the concept characterized by the finite automaton, given on each trial only the information of whether or not his classification of a given stimulus display is correct. I have mentioned line drawings here because I do not want to concentrate entirely on language, but it would also be possible to think of the learning in terms of recognizing grammatical strings of some regular language. Because I do not like to think of language learning wholly in terms of grammar, I prefer in the present context a geometrical example. Let us suppose that each line drawing consists of a finite number of line segments. A typical example might be a line drawing consisting of three segments forming a triangle but with one of the line segments, and only one, extending beyond the triangle. On each trial the learner is asked to say whether the particular line drawing shown is an instance of the display and after he has given this information he is told simply that his answer is correct or incorrect.

What the theory postulates is a sequence of implicit responses or, if you prefer, internal responses by the learner prior to giving the answer of 'yes' or 'no' to classify the display. Informally it is assumed that the implicit or internal responses are not available for observation and cannot be directly reinforced. The theory does not require this as an assumption, but it is implicit in any experimental application of the theory. Reinforcement takes place not after each internal response occurs, which is considered a subtrial, but only after a trial consisting of a sequence of subtrials.

In other words, putting the matter in standard experimental terms, a subtrial corresponds to what we usually think of as a trial, but no reinforcement or conditioning takes place and we cannot observe the response that was actually made. Conditioning occurs only after a sequence of subtrials, and the whole sequence of subtrials is called a trial. In automaton terms, a subtrial corresponds to an automaton making one transition, that is, from one internal state to another, and a trial to processing an entire tape or input string.

The characterization of the theory requires seven primitive concepts. To begin with, there is the set S of stimuli and the set R of responses. The set E of reinforcements contains only two elements, e_1 and e_2; e_1 is the positive reinforcer, e_2 the negative one. In the interpretation intended here, e_1 is the giving of information to the learner that his response has been correct and e_2 is the giving of information that the response is incorrect. It is clear that exactly how this information is given can vary widely from one experiment to another and there is a variety of procedures for doing so. What is important is that the reinforcers have this information interpretation. The fifth primitive concept is a measure μ of saliency defined on the set of stimuli. I will not have much more to say about the saliency of stimuli, and exactly how it is handled is not too important in the present context. In many simple experiments it is often taken to be the cardinality of the number of stimuli presented. The concept of subtrial requires the introduction of M, which is a sequence of positive integers m_n. Each m_n indicates the number of subtrials on trial n. This notion is necessary to define the next primitive concept, the sixth one, that of the sample space X. Each element of X represents a possible experiment, i.e., an infinite sequence of trials where each trial n has m_n subtrials. Each trial is an $(m_n + 2)$-tuple consisting of three things: (a) the conditioning function at the beginning of the trial which is a partial function from S into R, where $C(\sigma) = r$ means that stimulus σ is conditioned to response r and $C(\sigma)$ undefined means that σ is unconditioned; (b) m_n triples of the form (T, s, r) where T is the set of presented stimuli, s is the set of sampled stimuli, and r is the response on a subtrial; and (c) the reinforcement which occurred at the end of the trial. The final primitive concept is the probability measure P on the appropriate algebra of events (subsets) of X. All probabilities are defined in terms of P. For simplicity of formulation, in the following axioms it is assumed that all events on which probabilities are conditioned have positive probability. There are three kinds of axioms: sampling axioms, conditioning axioms, and response axioms. The nondeterminate reinforcement is especially relevant to the conditioning axioms.

Sampling Axioms

(S1) *On every subtrial a set of stimuli of positive measure is sampled with probability 1.*

(S2) *If the same presentation set occurs on two different subtrials, then the probability of a given sample is independent of the subtrial number.*

(S3) *Samples of equal measure that are subsets of the presentation set have an equal probability of being sampled on a given subtrial.*

(S4) *The probability of a particular sample on trial n, subtrial m, given the presentation set of stimuli, is independent of any preceding subsequence of events.*

Conditioning Axioms

(C1) *On every trial with probability 1 each stimulus element is conditioned to at most one response.*

(C2) *If* e_1 *occurs on trial n, the probability is c that any previously unconditioned stimulus sampled on a subtrial will become conditioned to the response given on that subtrial, and this probability is independent of the particular subtrial and any preceding subsequence of events.*

(C3) *If* e_1 *occurs on trial n, the probability is 0 that any previously unconditioned stimulus sampled on a subtrial will become conditioned to a response different from the one given on that subtrial, and this probability is independent of the particular subtrial and any preceding subsequence of events.*

(C4) *If* e_1 *occurs on trial n, the conditioning of previously conditioned sampled states remains unchanged.*

(C5) *If* e_2 *occurs on trial n, the probability is 0 that a previously unconditioned stimulus sampled on a subtrial will become conditioned.*

(C6) *If* e_2 *occurs on trial n, the probability is d that any previously conditioned stimulus sampled on a subtrial will become unconditioned, and this probability is independent of the particular subtrial and any preceding subsequence of events.*

(C7) *With probability 1, the conditioning of unsampled stimuli does not change.*

Response Axioms

(R1) *If at least one sampled stimulus is conditioned to some response, then the probability of any response is the ratio of the measure of sampled stimuli conditioned to this response to the measure of all the sampled conditioned stimuli, and this probability is independent of any preceding subsequence of events.*

(R2) *If no sampled stimulus is conditioned to any response, then the probability of any response r is a constant guessing probability ρ_r, that is independent of n and any preceding subsequence of events.*

Note that the conditioning method used is simple. Conditioning occurs on trials that have a correct response, and deconditioning occurs on trials that have an incorrect response. Thus learning occurs on all trials, regardless of whether the response is correct or not. On the basis of these axioms, the following theorem, which represents a considerable improvement of the basic theorem in my 1969 article, can be proved. The improvement is due to the weakening of the methods of reinforcement.

THEOREM. *If $\mathcal{D}$ is any set of perceptual displays and G is a subset of $\mathcal{D}$ that can be recognized by a finite automaton, then there is a stimulus-response model that can also learn to recognize G, with performance at asymptote matching that of the automaton.*

One important point to note is that with nondeterminate reinforcement the theorem is, as one would expect, weaker. In the 1969 article the stimulus-response model at asymptote became isomorphic to the given finite automaton. In the present case, the result can only be one of behavioral equivalence or, in the ordinary language of automaton theory, the result is one of weak equivalence. On the other hand, it is exactly the result of weak equivalence as opposed to isomorphism that is characteristic of neobehaviorism.

It is clear that the nondeterminate reinforcement used in the theory I have just formulated is about the weakest version of reinforcement that is interesting, with the possible exception of giving only partial reinforcement, that is, reinforcement on certain trials. In actual learning, for example, in the learning of mathematics or in the learning of language, there are many situations in which much more decisive and informative reinforcement is given. It is not difficult to show that the more determinate the reinforcement, the faster learning will be in general for organisms of a given capacity. In the long and tangled history of the concept of reinforcement it has not been sufficiently emphasized that reinforcement is delivery of information, and a particular structure of information is implicit in any particular scheme of reinforcement. An exhausting but not exhaustive analysis of different structures of reinforcement is to be found in Jamison *et al.* (1970). (So many detailed theoretical computations were made in this article that it has scarcely been read by anyone; it does provide a good sense of how complex things rapidly become when reinforcement schemes that have even mildly complex information structures are used.)

It is important for the present discussion to consider one of the weakest structures of nondeterminate reinforcement and to emphasize the point that it is the nondeterminate character of the reinforcement that moves us out of the arena of classical observability of responses and permits the introduction of a repertoire of implicit or internal responses that are not in general observable. The reinforcement does not create the implicit responses, but when we have determinate reinforcement the theory is not applicable to situations in which implicit or internal responses occur.

It is also important to note that it is really a matter of terminology and not of substantive theory whether these implicit responses are called responses as such or are called internal states. It would be easy enough to reformulate the axioms given above and to replace responses with internal states except for the response that occurs at the end of a trial. This terminological change would not affect the axioms in any way and might be a useful change for the purposes of emphasizing the move from behaviorism to neobehaviorism.

It is worth mentioning that the implicit responses that we might want to baptize as internal states are often observed as taking place even when we do not know their exact form. A good example occurs in the case of subvocalized articulatory responses that are characteristic of most silent adult readers. Self-awareness of such subvocal responses is unusual, and I hasten to add that it is not possible to 'read off' from the subvocal responses the words being read. In any case, to keep the behaviorist flavor of neobehaviorism I shall continue to talk about implicit responses rather than internal states or at least will not restrict myself to the terminology of internal states.

3. LEARNING PARTIAL RECURSIVE FUNCTIONS

Even in the case of nondeterminate reinforcement it is not difficult to extend the learning theorems for stimulus-response models beyond the theory of finite automata. We can, in fact, set our sights on the full set of computable objects, of what is known in the literature as the set of partial recursive functions. There are many equivalent definitions of this set of functions: functions computable by a universal Turing machine, partial recursive functions defined by partial recursive schemata, functions that are λ-definable, functions that satisfy a normal algorithm in the sense of Markov, and so on. Because of the extensional equivalence of all these definitions, there is good general agreement that the intuitive notion of computable can be characterized in a number of different ways, all of which are intuitively correct.

Classically, the computable functions, or partial recursive functions, have been defined as arithmetic functions from n-tuples of natural numbers to natural numbers, but it is easy to consider instead the partial recursive functions defined for a fixed finite vocabulary, and thus to consider functions that are more closely related to problems of language learning.

More than ten years ago, Shepherdson and Sturgis (1963) showed that one can use a quite simple set of instructions for register machines to write programs to compute any partial recursive function over a fixed finite vocabulary.

Let me recall how simple a register machine is. All we have is a potentially infinite list or sequence of registers, but any given program uses only a finite number. Exactly three simple kinds of instructions are required for each register. The first is to place any element of the finite vocabulary at the top of the content of register n; the second is to delete the bottommost letter of the content of register n if the register is nonempty; because any computation takes place in a finite number of steps, the content of any register must always be finite in length. The third instruction is a jump instruction to another line of the program, if the content of register n is such that the bottommost or beginning letter is a_i; in other words, this is a conditional jump instruction. Thus, if we think of the contents of registers as being strings reading from left to right we can also describe the instructions as placing new symbols on the right, deleting old symbols on the left, and using a conditional jump instruction in the program when required.

It is straightforward to give a formal definition of programs for such an unlimited register machine, but I shall not do so here; it is clear that a program is simply made up of lines of instructions of the sort just described. The important point is that it may be proved that given any partial recursive function computable over a finite vocabulary then a program that computes exactly that function for any given input string can be written in terms of the instructions stated above. The potentially infinite memory of an unlimited register machine both in terms of the number of registers and of the size of each register is a natural mathematical idealization. It is also possible to define a single register machine with instructions of the kind just stated and to show that a single register is also adequate. The use of such a single register moves the concept of a register machine close to that of a Turing machine.

In the present context the details are not important. What does seem intuitively desirable is the move from abstract machines with abstract internal states to programs written in terms of instructions each of which has an intuitive meaning. It is much easier to think about particular problems, especially problems of some complexity, in this fashion. Looked at in this way, the learning problem becomes one of writing internal programs to produce appropriate outputs for any given input. Thus, in the case of the kind of geometric problem discussed earlier, the input would again be a description in a finite vocabulary of a line drawing and the output of the program should be a classification response.

In extending learning theory to arbitrary partial recursive functions and using the concept of unlimited register machine just described, there are certain difficulties we have to be careful of theoretically. It would be easy to set the problem up so that, with the infinite set of registers and the unbounded length of possible programs, we would not get asymptotic convergence. There are various ways to bound the situation or to arrange the order of development so as to get the appropriate asymptotic theorem. I shall not enter into details, because it seems to me that there is a problem of an entirely different sort that is more critical and needs to be brought under scrutiny.

The weakness of asymptotic theorems of the kind given in the preceding section and of the kind hinted at for the present context of partial recursive functions is exactly the weakness of the theory of partial recursive functions itself as a theory of actual computers. Ignoring learning, for example, we may want to say that because we have a theory of computable functions we have a theory of computers. Nothing could be much further from the case.

For people involved in actual complex problems of programming, this seems like a ludicrous claim. The reason is that there is an enormous gap between the theory of computable functions and most of the questions we want to answer about computers. To a large extent the same thing is true about learning. It has sometimes been the claim of psycholinguists that in principle no stimulus-response theory could give an account of language learning. This claim is false, but proving it false is not nearly as important as developing an adequate positive theory.

There are many ways of saying what the important initial features of the positive theory should be. One way of formulating the matter is that we should be able to compute approximately the expected trial of last error for a given problem and for a learner with a given history of mastery of previous problems. If the theory is to match human performance, then the expected trial of last error must match approximately the performance of humans. If the theory is one for computer learning, then to make the situation interesting the expected trial of last error must be small enough to be testable for problems that seem appropriately simple.

If we find, for example, that the theoretical expected trial of last error is orders of magnitude larger than what we obtain in actual human learning or what we might expect or hope to obtain in computer learning, to take simply these two classes of learners, then clearly the theory is not yet thoroughly worked out as a satisfactory empirical theory to provide a theoretical basis for neobehaviorism.

Cognitive psychologists are properly impatient if all that can be offered them is the kind of general theory I have sketched. It is my view that the approach of cognitive psychologists or of psychologists interested in complex problem solving or information processing (Newell and Simon, 1972, is a good example) could be fit within a neobehaviorist framework if a proper amount of structure is assumed and not mastered from scratch. Cognitive psychologists are interested in studying complex problem solving or complex aspects of memory, for example. In general they are currently not very much interested in learning in the fundamental sense characteristic of the kind of theory I am describing in this article. There is not a formal inconsistency in the two viewpoints. There is currently a focus on different matters, but I think it is important for the future that a stronger convergence between the two viewpoints be attempted.

If we begin from the kind of neobehavioristic theory of learning I am sketching, then the general line of how to reach that convergence is clear in broad outline, and I turn to the analysis of that problem in the next section.

4. CHOOSING THE HIERARCHY OF PROBLEMS

The theoretical results for the learning of partial recursive functions discussed in the last section show that with extremely meager apparatus we can, even with nondeterminate reinforcement procedures, ultimately learn how to compute such functions, or in the terminology of the section before that, can learn to recognize a class of objects recognizable by a finite automaton. Humans, of course, are already endowed with a very rich structure for learning, but the theory emphasizes that this rich structure need not be there a priori. If, for example, one has in mind the development of a theory of learning for computers, then one might in principle not want to assume much structure at the beginning.

It is also easy enough to place human learners in a situation that is in practical terms too difficult for them. No one, for example, would consider teaching advanced mathematics to a child or an adult who did not have an appropriate prior background in mathematics. Many complex technical skills that take a long time to learn have a similar character. One would, for example, not hire an untrained person, no matter how bright and experienced in other areas, to do under a definite time pressure a large set of architectural drawings.

In every area of specialized knowledge or skill the learner is expected to work his way through a hierarchy of problems. Although the argument is usually not put in explicit terms, it is understood that the external organization of such a hierarchy is essential for almost all learners to make any reasonable progress.

Organizing the hierarchy of tasks or problems that the learner must master is a typical traditional problem in the organization of curriculum. A great deal of practical experience and wisdom is embodied in the major areas of curriculum in the schools and universities, but the theory of such matters is still in quite an elementary state.

Perhaps the central reason for the elementary state of the theory is that, in regard to the deeper principles for organizing the hierarchy of concepts or skills to be learned, we have as yet rather poor ideas of how the hierarchy is internally absorbed by the learner. It is probably agreed by everyone that development of an internal hierarchy is as essential as the presentation of problems in an external hierarchy.

From the standpoint of asymptotic arguments, neither the internal nor the external hierarchy is required, but once we turn from asymptotic questions to questions of actual learning and concern with the efficiency of learning then it is obvious that detailed attention must be given to both internal and external hierarchies.

It seems to me also important to recognize that to a reasonable degree the internal hierarchy is as much subject to control and variation as is the external hierarchy.

Certainly we are not conscious of how our memories work, but we are conscious of various ways of improving memories or facilitating the ways in which we remember things. A good example would be the traditional method of associating memories with places as a method of facilitation.

I take it that a central goal of cognitive psychology is to characterize the variety of internal structures and their functions. As already remarked, it is characteristic of contemporary cognitive psychology to be not much concerned with the kind of internal hierarchies that can be learned but rather to study that which is already there on the basis of prior experience and learning, but not prior experience and learning that has itself been a subject of experimental or analytical study. I see the convergence of cognitive psychology and the neobehaviorist kind of learning theory I have been sketching in the study of the kinds of internal hierarchies that can be learned and that will prove useful to learn in order to master complex concepts and skills.

An example with considerable practical implications is the learning of skills for giving mathematical proofs. Students all over the world are taught the elements of proof-making skills, as we might call them, but the psychological study of mathematical proofs is as yet in its infancy. Neither the theory of learning nor the current theories of cognitive psychology has yet much to offer to provide a deeper insight into the development or use of proof-making skills. As far as I know, there is not one single psychological study of a systematic kind about mathematical proofs at the level of difficulty, say, of a first-year graduate course in mathematics.

For definiteness let us return to the classification of line drawings mentioned earlier. As each new problem, that is, each new classification, is learned, the appropriate internal structure is that a subroutine is added to the programs being written by the internal register machine. These new subroutines are named; indeed, they might well be named with the appropriate English words. Moreover, a new predicate is added to the internal language for scanning objects and this new predicate will, if things work out right, be used in the near future in the analysis of new classes of problems. I am under no illusion that the creation of subroutines that can be called and the creation of new predicates for approximate classification of objects when faced with new problems constitute a sufficient apparatus. Not only will additional structures be needed, but the articulation of relations between the structures is at least as important and as delicate a problem. My only point in the present discussion is that it is my firm conviction that when we talk about learning beginning from scratch, the hierarchy of problems solved will itself have a determining effect on the creation of the internal hierarchy that will be used in solving subsequent problems. In advanced and difficult areas of problem solving the exact hierarchy that is internalized probably has a great deal to do with the ability to solve new problems. In areas of science that have received a considerable development, breaking through the highly developed current hierarchy of concepts and ideas is often the most important single step in solving an open problem.

I believe that computer-learning experiments can play the same insightful role in understanding how an internal hierarchy is created as have the experiments of biochemists with the chemical conditions for creating the first lifelike molecules (Calvin, 1975). One does not expect in these biochemical experiments to move at any rapid pace from the creation of lifelike molecules to the synthetic creation of complicated living organisms. It is the objective of such investigations to gain fundamental insight into how the complex molecules essential to our forms of life evolved from much simpler structures. In the same way, experiments on computer learning provide an opportunity to gain insight into the way that internal hierarchies are formed in the solving of a hierarchy of problems. Such experiments have as yet only begun, because we still need the detailed articulation of a learning theory that will make such matters practical, but there is currently a great deal of work going on relevant to these matters in many different intellectual centers throughout the world and I am mildly optimistic about progress in the near future.

5. PROBLEMS OF PREDICTION

I find myself very much in agreement with what Nelson has to say about psychology being possibly quite a different subject from physics. We can expect to need a theory of individual prediction rather than a theory of how most organisms work most of the time in the same way.

I shall not try to summarize his arguments but to state my own views from a slightly different viewpoint. From inspection of computer hardware it is clearly ludicrous to think that one can predict the kinds of computer programs that will be written for the computer system. Of course, certain very gross and uninteresting statements can be made, but statements that predict in any detail the actual programs that will be written are obviously out of the question. In ordinary scientific terms, knowledge of the hardware in no sense determines knowledge of the software. It seems to me that there is good evidence that the same situation is approximately true for human beings. Knowledge of how the physical hardware of the brain works will not necessarily tell us very much at all about the psychological aspects of human activities, especially the more complex ones. For example, it is nice to know that language activity is ordinarily centered in the left hemisphere of the brain, but it seems quite evident that in no foreseeable future will dissection of the left hemisphere of an unknown person be able to identify the language he actually spoke, whether it be English, Russian, Chinese, or what not. Location and identification of more particular skills or memories on the part of particular humans is clearly an even more impossible task. The software of the brain will not be reduced to the hardware in any way that seems feasible at the present time, and in this sense it seems to me a strong claim can be made that psychology is not going to be reduced to physiology and biology.

It is this line of argument that makes psychology as fundamental a science as physics. On various occasions mistaken views have been held about the reduction of psychology to physiology or, in even more bold terms, the reduction of psychology to physics. Nothing, it seems to me, is further from being the case, and it is because of this absence of any evidence that any reduction can take place that theses about behaviorism remain important. Psychological concepts, complex skills, and, in a still more traditional terminology, mental events as occurring at least in other persons and other animals can be known only from behavioristic evidence. We will not obtain that evidence from chemical or physical examination of the cells of the body. We will not obtain it by rationalistic methods of knowing. Behaviorism as a fundamental methodology of psychology is here to stay, but the room that it occupies is sufficiently large to admit a dazzling array of mental furniture. Clear recognition that there is mental furniture inside the room is why the sign over the door should now be changed from behaviorism to neobehaviorism.

25

LEARNING THEORY FOR PROBABILISTIC AUTOMATA AND REGISTER MACHINES, WITH APPLICATIONS TO EDUCATIONAL RESEARCH

The first part of this article reviews my work and that of my collaborators in the learning theory of probabilistic automata and register machines. The second part is concerned with specific applications to educational research, especially to the learning of elementary mathematics.

1. THEORY

Asymptotic theory. The current asymptotic theory is in reasonable shape. One can give an asymptotic theory having the following intuitive content. Given any classification problem that can be characterized by a finite automaton, there exists a stimulus-response model that under nondeterminate reinforcement will asymptotically be equivalent to the classification behavior of the automaton. What is important about this theorem is that

*Reprinted from *Structural models of thinking and learning* (ed. by H. Spada and W. F. Kempf), 1977, pp. 57–79. Bern: Hans Huber.

the reinforcement is nondeterminate, that is, the stimulus-response model can learn from only "yes-no" reinforcements regarding the correctness of classification. This theorem is in contrast to that obtained in Suppes (1969b), which depended upon determinate reinforcement. The earlier theorem is, of course, stronger, because with determinate reinforcement it is possible to obtain isomorphism asymptotically between the stimulus-response model and the given finite automaton. With nondeterminate reinforcement the best one can expect to get is behavioral equivalence. The theory of stimulus-response models with nondeterminate reinforcement developed by William Rottmayer and myself is outlined in the preceding article in this volume. The basic theorem is this.

THEOREM 1. *If θ is any set of perceptual displays and G is a subset of θ that can be recognized by a finite automaton, then there is a stimulus-response model that can also learn to recognize G, with performance at asymptote matching that of the automation.*

I now turn to a comparable development for register machines. In the formal characterization of such machines I follow Shepherdson and Sturgis (1963). First, let me recall how simple a classical register machine is. All we have is a potentially infinite list or sequence of registers, but any given program uses only a finite number. Exactly three simple kinds of instructions are required for each register. The first is to place any element of the finite vocabulary at the top of the content of register n; the second is to delete the bottommost letter of the content of register n if the register is nonempty; because any computation takes place in a finite number of steps, the content of any register must always be finite in length. The third instruction is a jump instruction to another line of the program, if the content of register n is such that the bottommost or beginning letter is a_i; in other words, this is a conditional jump instruction. Thus, if we think of the contents of registers as being strings reading from left to right we can also describe the instructions as placing new symbols on the right, deleting old symbols on the left, and using a conditional jump instruction in the program when required.

It is straightforward to give a formal definition of programs for such an unlimited register machine, but I delay this for the moment. It is clear that a program is simply made up of lines of instructions of the sort just described. The important point is that it may be proved that, given any partial recursive function computable over a finite vocabulary, a program that computes exactly that function for any given input string can be written in terms of the instructions stated above. The potentially infinite memory of an unlimited register machine both in terms of the number of registers and the size of each register is a natural mathematical idealization. It is also possible to define a single-register machine with instructions of the kind just stated and to show that a single register is also adequate. The use of such a single register moves the concept of a register machine close to that of a Turing machine.

Before looking at the details of register machines as we shall want to construct them for learning purposes, it will be useful to review the seven primitive concepts used in the nondeterminate stimulus-response theory formulated in the previous article. The set S of stimuli, the set R of responses, the set E of reinforcements, and the measure of saliency on the set of stimuli all remain unchanged, in principle if not in practice. The concept of a subtrial of the sample space Ω and the probability measure P will change somewhat, but only in their details. We must also introduce as a new primitive concept the set of internal instructions of the register machine, as well as the registers themselves. To make the theory as simple as possible in general formulation, I shall first assume that the set of registers is unbounded in number but that in a given program only a finite number are used, and then later assume only a fixed finite number is available at all. Still another general primitive concept is that of the internal language used for encoding stimulus displays. In the present formulation of the theory I shall, in fact, not enter into the relation between the set of stimuli and the encoding language but deal only with the already encoded representation of the display. This level of abstraction seems appropriate for the present discussion. It is a matter of a detailed theory of perception to work out the relationship between stimulus presentations and internal encodings. Thus, as the theory is presented here, the concept of stimulus is actually nonfunctional, but this is not because of any fundamental belief that it should be nonfunctional but only due to a drastic abstract simplification of the theory. I comment on this issue again in the discussion of applications in part II.

The concept of a program internally constructed replaces that of subtrial. For purposes of the general theory I shall leave matters as stated above; namely, there are three kinds of instructions and the input alphabet of the register machine is a finite nonempty set V.

To make matters more explicit and formal but without attempting a complete formalization, I introduce the following definitions. First, $\langle n \rangle$ is the content of register n before carrying out an instruction; $\langle n' \rangle$ is the content of register ,n after carrying out an instruction. Second, a register machine has 1) a denumerable sequence of registers numbered $1, 2, 3, \ldots$, each of which can store any finite sequence of symbols from the basic alphabet V, and 2) three basic kinds of instructions:

(a') $p_N^{(i)}(n)$: Place a_i on the right-hand end of $\langle n\rangle$.

(b') $D_N(n)$: Delete the left-most letter of $\langle n\rangle$ if $\langle n\rangle \neq 0$.

(c') $J_N^{(i)}(n)[E1]$: Jump to exit 1 if $\langle n\rangle$begins with a_i.

The notation $E1$ is a variable for the line number to jump to. If the jump is to a nonexistent line, then the machine stops. The parameter N shown as a subscript in the instructions refers to the set of registers left unchanged when the program is completed. (This point is made more explicitly in the definition given below.)

A *line* of a program of a register machine is either an ordered couple consisting of a natural number $n \geq 1$ (the line number) and one of the instructions (a) or (b), or an ordered triple consisting of a natural number $n \geq 1$, one of the instructions (c), and a natural number $m \geq 1$. The formal interpretation of this definition is obvious and will not be given.

A *program* (of a register machine) is a finite sequence of k lines such that 1) the first member of the i^{th} line is i, and 2) the numbers m that are third members of lines are such that $1 \leq m \leq k+1$. The parameter k is, of course, the number of lines of the program. I shall also refer to programs as *routines*. How a register machine *follows* a program or routine is intuitively obvious and will not be formally defined. *Subroutines* are defined like programs except 1) subroutines may have several exits, and 2) third members of triples may range over $E_1, \ldots, E_k$, these variables being assigned values in a given program.

I shall not give the formal definition of a partial recursive function defined over the alphabet V. It is any intuitively computable function. Given V, the finite vocabulary, then, as usual in such matters, V^* is the set of finite sequences of elements of V; in the present context, I shall call the elements of V^* 'codings'. Let f be a function of n arguments from $V^* \times \cdots \times V^*$ (n times) to V^*. The basic definition is that f is computable by a register machine if and only if for every set of natural numbers $\{x_1, \ldots, x_n, y, N\}$ with $y \neq x_i$ for $i = 1, \ldots, n$ and $x_1, \ldots, x_n, y \leq N$ there exists a routine $R_N(y = f(x_1, \ldots, x_n))$ such that if $\langle x_1\rangle, \ldots, \langle x_n\rangle$ are the initial contents of registers $x_1, \ldots, x_n$ then

1) if $f(\langle x_1\rangle, \ldots, \langle x_n\rangle)$ is undefined the machine will not stop,

2) if $f(\langle x_1\rangle, \ldots, \langle x_n\rangle)$ is defined, the machine will stop with $\langle y\rangle$, the final content of register y, equal to $f(\langle x_1\rangle, \ldots, \langle x_n\rangle)$, and with the final contents of all registers $1, 2, \ldots, N$, except y, the same as initially.

I turn now to the axioms for register-machine learning that roughly parallel those given in the previous article for stimulus-response models with nondeterminate reinforcement. Register 1 is reserved for the response. In the present case, if the register is cleared, the response is that the stimulus display is an instance of the concept in question and if the register is not empty, the answer is negative. Moreover, if the program stops before completion, the answer also is negative. The program constructed is stored in a *program stack*, which is just a special register designated for this purpose.

In addition to this notion of a register stack, the register machine will be restricted to a nonempty finite set M of registers, numbered $1, \dots, |M|$, where $|M|$ is the cardinality of M. Three other primitive concepts are needed: the function k from V^* to the real numbers—for each x in V^*, $k(x)$ is the upper bound on the running time or length of program for computing on x; the real number c is the parameter of the geometric distribution on the number of lines of program constructed; and the real number g is the parameter of the geometric distribution needed for conditional jump instructions (see Axiom I2 below). The concepts and axioms apply only to learning partial recursive functions of a single argument. Moreover, only classificatory functions are treated; in particular, functions that have only two values, 0 and 1. Thus the theory is restricted to parallel the earlier treatment of stimulus-response models. The technical details mentioned in the axioms and following theorem should be still more explicit and formal than they are, but because the asymptotic theory as such is of only limited practical interest, I have made the treatment rather brief.

Initial data axiom

(D1) At the start of each trial, there is an x in V^* such that $\langle 1 \rangle = x$.

Program construction axioms

(I1) If the program stack is nonempty, no new program is constructed (because the one already there will be used).

(I2) Given that the program stack is empty at the beginning of a trial:

1) the probability of constructing a program of n lines is $c(1-c)^{n-1}$ with $0 < c < 1$, independent of the trial number and any preceding subsequence of events;

2) given that a line is constructed, the probability of sampling an instruction is uniformly distributed over M, V, and the three types

of instruction, independent of the trial number and any preceding subsequence of events;

3) if a conditional jump instruction is sampled, the line (m) "jumped to" is sampled geometrically with parameter g, $0 < g < 1$, independent of the trial number and any preceding subsequence of events.

Program erasure axioms

(E1) If e_1 occurs at the end of a trial, the program stack remains unchanged.

(E2) If e_2 occurs at the end of a trial, the program stack is cleared.

Response axioms

(R1) If the program halts and $\langle 1 \rangle$ is empty, the response is "yes."

(R2) If the program halts and $\langle 1 \rangle$ is nonempty, the response is "no."

(R3) If the program does not halt by elapsed time $k(x)$ for input data x in V^*, the response is "no.", and the next trial is begun.

On the basis of these axioms we may prove an asymptotic theorem corresponding to Theorem 1 for stimulus-response models.

THEOREM 2. *Let f be any 0–1 partial recursive function of one argument over the alphabet V such that a program for f exists with running time less than $k(x)$ for x in V on register machine $\mathfrak{M} = (M, V, k, c, g)$. Then f is asymptotically learnable with probability one by $\mathfrak{M}$.*

Proof. Let $\mathcal{P}$ be a program for $\mathfrak{M}$ that computes f for argument x in V^* in running time less than $k(x)$. Let $C \subseteq V^*$ be the set of instances of the concepts and $\neg C$ its complement, i.e., $\neg C = V^* - C$. Then we shall impose as a stimulus-display sampling distribution

$$P(x \in C) = P(x \in \neg C) = \frac{1}{2}.$$

If C or $\neg C$ is finite, a uniform sampling distribution is used; if C or $\neg C$ is infinite, a geometrical distribution on the length—with uniform distribution on a given length—is used. (The exact nature of this distribution is unimportant.)

Let $\mathcal{P}'$ be a constructed program which is incorrect for at least one x in V^*. Then because P (sampling x) > 0, with probability one $\mathcal{P}'$ will be erased. On the other hand, the probability that $\mathcal{P}$ will be constructed on any trial for which the program stack is cleared is positive and independent of the trial number. Thus asymptotically $\mathcal{P}$ or an equivalent correct program will be constructed with probability one. Details are omitted but are similar to those given in Suppes (1969b).

Asymptotic theorems of the kind just proved are of limited application when no bounds on the expected last error are given, or when a lower bound is given that is far too high to account for any actual learning that is of interest.

The nonasymptotic role of hierarchies. In the remainder of this section I examine some of the consequences of building a hierarchy of internal subroutines to match, i.e., to solve, a hierarchy of external problems. As a first example, made much too simple in order to make some explicit computations, I consider a disjunctive concept made up of n disjoint cases. Only one register is required, the alphabet is the set $\{0, 1\}$, and there is no jump instruction, but only the four instructions for deleting letters on the left or adding them on the right. Let the program be at most 10 lines for each case. Then assuming a uniform distribution on sampling of instructions and of the number of lines (1 to 10), the probability of each program of at most 10 lines can be directly computed. More importantly in the present instance, we can easily compute the possible number of programs: 4 of length 1, 16 of length 2, and in general 4^n of length n, with $1 \leq n \leq 10$, for a total of $(4^{11} - 4)/3$, which is approximately 4^{10}. If now at the second stage programs are put together using only original instructions and the n subroutines from individual cases, with programs of length at most 2n permitted, then there are $[(n+4)^{2n+1} - (n+4)]/(n+3)$ possible programs, which is approximately $(n+4)^{2n}$. On the other hand, if a single program is developed in one step with $10n$ lines, the number of possible programs is approximately 4^{10n}. Consider, for example, the case $n = 3$. Then 4^{30} is order of magnitudes larger than $7^6 + 4^{10}$.

The details of this example are not important. I have not attempted to fix them sufficiently to determine in each of the two approaches the number of the possible programs that are correct. Ordinarily in both the hierarchical and nonhierarchical approach this number would be a very small percentage of the total. The gain from the hierarchical approach is evident enough already in this example. But even a simple hierarchical approach seems to lead to learning that is too slow. What this suggests is that the degree of hierarchical structure must be even more extensive.

There are two additional important conceptual matters that should be enlarged upon, independent of any particular applications.

The first is that the learning theory outlined here is not in any sense restricted to the learning of algorithms, as Theorem 1 about concepts recognizable by finite automata might suggest. Theorem 2 already in principle applies to concepts whose extensions are recursively enumerable but not recursive, and thus not algorithmic in character. The probabilistic approach to learning or problem-solving characteristic of the theory applies in principle just as well to tasks that do not have any obvious algorithmic characterization. The learning of strategies for finding mathematical proofs is an example of considerable pedagogical importance. Note that even in a domain for which an "obvious" solution algorithm exists, it may be totally impractical to use it, and thus a "creative" approach is needed. A good example would be proofs of theorems in elementary algebra or elementary Euclidean geometry for which there exists an algorithmic decision procedure by well-known results of Tarski (1951), but for good theoretical reasons there is no hope of a general application of his procedures, because it may be shown that the lower bound on the number of steps, as a function of the length of a formula, is in general large enough to make practical use of the decision procedure out of the question.

Consequently, future students as well as present ones will need to continue to learn how to solve elementary algebraic and geometrical problems in a nonalgorithmic, creative fashion. On the other hand, the learning goals we set for students differ widely when the task, or set of tasks, is algorithmically solvable in a practical manner and when it is not. Thus we expect students to be able to solve correctly an indefinitely large number of algorithmic arithmetic exercises, but we have much more limited expectations for their "proof-making" skills.

The other conceptual matter is the clear recognition of the great gains to be obtained from using methods of reinforcement that are stronger than the weakest nondeterminate ones. This fact is recognized in all ordinary teaching. It is the purpose of explanations, of didactic lectures, of verbal corrections of students, of attempts at explaining to them why they have made a mistake, and of a variety of other diagnostic approaches that lead to explicit verbal communication to students. I have discussed earlier the reasons for not going all the way in the other direction to completely determinate reinforcement. In this case, we end up with methods that are too strong and that cannot be practically realized.

The conceptual subtlety of the reinforcement-information problem is the difficulty of giving an explicit account of how the student processes the complex verbal information that is given to him. It is easy enough to have a theory of that when the information given to him is simply the weak nondeterminate reinforcement concerning the correctness or incorrectness of his answer. When a complex verbal description of his difficulty or of his mistake is communicated, the understanding of the process that is taking place is another matter. It is clear that the framework I have presented here, although in principle it may be strong enough to account for such complex communications to the student, in practice is much too undeveloped to do so. The full complexities of the theory of natural-language processing required to take account of these matters is awe-inspiring, but we can make various approximations and I attempt to do that in the next part of this paper dealing with some applications. I want to close this theoretical section, however, with a statement on the great importance of developing the systematic theory of such complex information transfer in the form of semideterminate reinforcement from instructor to student. Until we have a deeper theory of the natural-language processing involved, we shall, I fear, be inevitably limited in our theory of instruction.

2. APPLICATIONS

In this part I examine ways in which probabilistic automata and register machines can be applied to specific problems of educational research. I begin with a survey of the empirical and theoretical results that have been obtained thus far. On the one hand, the theory is fairly well developed and has been tested in simple cases rather extensively. Moreover, the theory has a certain fundamental property I consider essential: The theory is rich enough to process the problems of elementary arithmetic beginning with a schematic form of visual perception. It is my view that any serious theory of elementary mathematics learning at the present time should satisfy such a minimal requirement of processing. On the other hand, it is also important to emphasize how schematic the present theory is. The conception, for example, of visual perception and the processing of visual stimuli by students is extremely rudimentary. This also applies to the processing of auditory stimuli.

The theoretical and empirical work on probabilistic automata is to be found in Suppes (1969b), Suppes and Morningstar (1972), and Suppes (1973). I shall not survey the details here, but the theoretical objective of these studies is to introduce meaningful probabilistic parameters for state transitions for probabilistic automata that are informationally adequate to the algorithms of elementary arithmetic. The empirical objective is to use data to estimate the numerical values of the parameters and to test the goodness of fit of the models with respect to the data. The analysis given in the publications referred to deals mainly with performance data and not with learning. It is especially for this reason that a review of the results obtained has been omitted.

I turn now to register machines. The theoretical approach is to assume that the structure of the student for the learning of elementary arithmetic may be represented by a register machine with a small finite number of registers and a small number of elementary instructions. Using the small number of registers, which are distinguished on psychological grounds as registers with stimulus support and those without, algorithms for solving elementary arithmetic problems can be built up, and the realism of these algorithms in relation to the actual learning and performance of students can be studied. The central problem, of course, is to give a reasonable account of the kinds of errors that students make.

To provide a concrete illustration of a register machine for elementary mathematics, I characterize here in schematic form a register machine adequate for column addition and similar tasks. For column addition, three registers suffice in our scheme of analysis. First there is the stimulus-supported register (SS) that holds an encoded representation of a printed symbol to which the student is perceptually attending. In the present case the alphabet of such symbols consists of the 10 digits and the underline symbol ‘___’. As a new symbol is attended to, previously stored symbols are lost unless they are transferred to a non-stimulus-supported register. The second register is the non-stimulus-supported register (NSS). It provides long-term storage for computational results. The third register is the operations register (OP) that acts as a short-term store, both for encodings of external stimuli and for results of calculations carried out on the contents of other registers. It is also primarily non-stimulus-supported.

It is important to note that in the case of the algorithms of elementary mathematics the number of registers is quite small and the amount that the student is expected to hold in a register is also severely restricted. In contrast to the way computers are expected to perform algorithms, the student makes extensive use of stimulus-supported registers and is able continually to refresh by perception his memory of the main data in front of him. In many respects this is the most single striking conceptual difference between the way human beings perform elementary mathematical algorithms and the way they are performed in electronic computers.

We drastically simplify the perceptual situation by conceiving each exercise as being presented on a grid with at most one symbol in each square of the grid. For column addition we number the coordinates of the grid from the upper right-hand corner. Thus, in the exercise

$$\begin{array}{rr} & 18 \\ & 32 \\ + & 46 \\ \hline \end{array}$$

the coordinates of the digit 8 are (1,1), the coordinates of 2 are (2,1), the coordinates of 6 are (3,1), the coordinates of 1 are (1,2), and so forth, with the first coordinate being the row number and the second being the column number.

The restricted set of instructions we need for column addition are the following 10.

Attend (a, b): Direct attention to grid position (a, b).

$(+a, +b)$: Shift attention on the grid by $(+a, +b)$.

Read in (SS): Read into the stimulus-supported register the physical symbol in the grid position addressed by Attend.

Lookup
(R1) + (R2): Look up table of basic addition facts for adding contents of register (R1) and (R2) and store the result in (R).

Copy
(R1) in (R2): Copy the content of register (R1) in register (R2).

Deleteright (R): Delete the rightmost symbol of register (R).

Jump L: Jump to line labeled L.

Jump (val)R,L: Jump to line labeled L if content of register (R) is val.

Outright (R): Write (output) the rightmost symbol of register (R) at grid position addressed by Attend.

End: Terminate processing of current exercise.

Exit: Terminate subroutine processing and return to next line of main program.

Of the 10 instructions, only *Lookup* does not have an elementary character. In the complete analysis it has the status of a subroutine built up from more primitive operations such as those of counting. It is, of course, more than a problem of constructing the table of basic addition facts from counting subroutines; it is also a matter of being able to add a single digit to any number stored in the non-stimulus-supported register (NSS) or (OP), as, for example, in adding many rows of digits in a given column. I omit the details of building up this subroutine.

It should also be obvious that the remaining nine instructions are not a minimal set; for example, the unconditional jump instruction is easily eliminated.

To illustrate in a simple way the use of subroutines, I consider the simple one for outputting all the digits in a register with, of course, the outputting of the digits from right to left as in the standard algorithm of column addition.

```
Output (R)
Put         Outright (R)
            Deleteright (R)
            Attend (0,+1)
            Jump (Blank) R, Fin
            Jump Put
Fin         Exit
```

I turn now to problems of learning, and, as in the case of the analyses in Suppes (1973), I restrict myself to the case of single-column addition, but with an indefinite number of rows. This means that in general the output subroutine just described will need to be used.

Let me reproduce here the internal program shown on the left and the verbal instructions used for instruction on the right. This material is shown in Figure 1.

	Internal Program		*Verbal Instructions*
	Attend (1,1) Readin	c_1	Start here (pointing)
   Opr.	Transfer (SS) to (OP) Attend(+1,0) Readin Lookup (OP)+(SS)	c_2	Add first two digits (pointing)
	Attend (+1,0) Readin Jump (0-9) SS, Opr	c_3	Now add again (pointing) (if conditional jump satisfied) or Notice end of column (pointing at __) (if conditional jump not satisfied)
	Attend (+1,0) Output (OP) End	c_4	Write answer here (pointing)

Figure 1. Single-column addition.

In Figure 1, learning parameters c_1, c_2, c_3 and c_4 are shown for the four segments of the program. These learning parameters have an abstract quality not directly related to the detailed axioms for program construction given in the preceding section. They permit us to develop at an abstract level simple learning models familiar from the literature of mathematical psychology. The simplest such model is the one that assumes independence of the four parts. If we treat the probability of successive errors combining to yield a correct response as having probability zero, then the mean probability for a correct response on trial n for the independence model is simply:

$$P_n \text{ (Correct Response) } = \Pi_{i=1}^{4}(1-(1-c_i)^{n-1}).$$

At the other extreme, a hierarchical model, also at the same general level of abstraction, postulates that the i^{th} segment of the program cannot be learned until the $i - 1^{st}$ segment is learned. This simple abstract hierarchical model leads to the following transition matrix, where state 0 represents all segments as unlearned, state 1 represents the first segment only as learned, etc.

$$\begin{array}{c|ccccc}
 & 4 & 3 & 2 & 1 & 0 \\
\hline
4 & 1 & 0 & 0 & 0 & 0 \\
3 & c_4 & 1-c_4 & 0 & 0 & 0 \\
2 & 0 & c_3 & 1-c_3 & 0 & 0 \\
1 & 0 & 0 & c_2 & 1-c_2 & 0 \\
0 & 0 & 0 & 0 & c_1 & 1-c_1
\end{array}$$

It is clear that neither one of these simple models, the independence model or the hierarchical model, gives an informationally adequate account of what is taking place in the sense characterized earlier.

Let us therefore examine to what extent a more adequate detailed model can be developed for the simple problem of column addition. The difficulties we face in doing so is evidence of the general difficulty of developing informationally adequate learning models that can at the same time be systematically compared to quantitative data.

Consider the learning parameter c_2. Suppose the possible instructions for the subroutine "Add first two digits" as shown in Figure 1 are sampled independently and with equal probability. These two simplifying assumptions make explicit computations manageable. Suppose further, and still more unrealistically, that the subroutine will be exactly four lines and will be constructed by sampling only the four instructions actually used. Then if the sampling is with replacement, c_2 is approximately .0039, which seems unrealistically small. If eight of the instructions are available, with (equal) probability .125 of being sampled with replacement, then $c_2 = .00024$. (These calculations assume there is exactly one correct program of four lines.) Almost any data taken from actual student performance will show that either of these estimates of c_2 is far too small.

The moral of this tally is that learning proceeds by smaller steps than naive intuition is inclined to surmise. The schematic computations just made raise interesting problems about past controversies in learning concerning the all-or-none versus incremental character of simple concept formation. Although most of the simple concept-formation experiments I have in mind are at heart algorithmic in character, the concept is not explicitly taught in an algorithmic fashion. It is characteristic of the experimental instructions not even to give a hint of there being an algorithm for identifying the presence or absence of the property exemplifying the concept. A number of experiments that support the all-or-none hypothesis are reported in Suppes (1965). I consider here the simple experiment in which five-year-old children were asked to discriminate between line drawings of triangles, quadrilaterals, and pentagons. The algorithm, obviously, is simply to count the number of sides, but the algorithm was not explicitly taught to the children as part of the experiment and they were not asked to verbalize their procedure at any stage. Because they found triangles easy to discriminate from quadrilaterals and pentagons but found it a good deal more difficult to separate quadrilaterals from pentagons, there is rather good evidence that they were not using a counting algorithm but rather a perceptual response of a different character. The strong support for all-or-none learning in this experiment does, however, indicate that the instructions for the experiment essentially put the children in a "frame" (to use the term made popular by Marvin Minsky in artificial intelligence). In the language being developed in this paper, being put within a particular frame would mean that most of the subroutines needed for giving the correct response were already called up and put in place by the experimental instructions. On this view, all-or-none learning is obtained because only a single instruction or, at most, only a very small number need to be added to the frame in order to give the correct response.

In experiments either with children or with adult human subjects, the instruction with which the experiment begins has the effect of a subject in the experiment constructing a frame within which he proceeds to work. Of course, in many cases the subject constructs a frame that has to be modified because he has a misunderstanding of what the experiment is about. It is considered the mark of a good experiment that the subject understands the instructions and this means that the general frame he establishes is the intended one, or nearly so, very early in the experiment—

I refer here of course to "direct" cognitive experiments, not to the kind of "misleading" frameworks intentionally established in many experiments in social psychology.

In the case of instruction in the schools, good organization of the curriculum as it is presented to students should lead naturally from one frame to another, so that the additional subroutines that must be organized are small in number as the student undertakes to master a new concept or skill. The kind of elementary computations exhibited in this paper tend to show why the steps in moving from one part of the curriculum to the next must be small and well organized in order for the curriculum to be successful with most of the students. This careful articulation of the curriculum is as important in the teaching of mathematics and science at the university level as at the beginning school level. It is in fact unfortunate that we have as yet very unsatisfactory empirical traditions of analyzing the learning that is involved in mastering more advanced topics in mathematics and science. The psychological investigation of these matters is almost untouched.

26

IS VISUAL SPACE EUCLIDEAN?

Philosophers of past times have claimed that the answer to the question, Is visual space Euclidean?, can be answered by *a priori* or purely philosophical methods. Today such a view is presumably held only in remote philosophical backwaters. It would be generally agreed that one way or another the answer is surely empirical, but the answer might be empirical for indirect reasons. It could be decided by physical arguments that physical space is Euclidean and then by conceptual arguments about perception that necessarily the visual space must be Euclidean. To some extent this must be the view of many laymen who accept that to a high degree of approximation physical space is Euclidean, and therefore automatically hold the view that visual space is Euclidean.

I begin with the question, How do we test the proposition that visual space is Euclidean? The first section is devoted to this problem of methodology. The second section provides a brief overview of the hierarchy of geometries relevant to visual phenomena. The third section reviews a number of answers that have been given to the question of the Euclidean character of visual space. I examine both philosophical and psychological claims. The final section is devoted to central issues raised by the variety of answers that have been given.

1. HOW TO APPROACH THE QUESTION

What would seem to be, in many ways, the most natural mathematical approach to the question has also been the method most used experimen-

*Reprinted from *Synthese*, **35** (1977), 397-421.

tally. It consists of considering a finite set of points. Experimentally, the points are approximated by small point sources of light of low illumination intensity, displayed in a darkened room. The intuitive idea of the setting is to make only a finite number of point-light sources visible and to make these light sources of sufficiently low intensity to exclude illumination of the surroundings. The second step is to ask the person making visual judgments to state whether certain geometrical relations hold between the points. For example, are points a and b the same distance from each other as points c and d? (Hereafter in this discussion I shall refer to points but it should be understood that I have in mind the physical realization in terms of point-light sources.) Another kind of question might be, Is the angle formed by points $a\ b\ c$ congruent or equal in measure to the angle formed by points $d\ e\ f$?

Another approach to such judgments is not to ask whether given points have a certain relation but rather to permit the individual making the judgments to manipulate some of the points. For example, first fix points a, b and c and then ask him to adjust d so that the distance between c and d is the same as the distance between a and b. Although the formulation I am giving of these questions sounds as if they might be metric in character, they are ordinarily of a qualitative nature—for example, that of congruence of segments, which I formulated as same distance. No metric requirements are imposed upon the individuals making such judgments. For instance, no one would naturally ask subjects in the experiments relevant to our question to set the distance between two points to be approximately 1.3 meters or to determine an angle of, say, 21 degrees.

Once such judgments are obtained, whether on the basis of fixed relations or by adjusting the position of points, the formal or mathematical question to ask is whether the finite relational structure can be embedded in a two- or three-dimensional Euclidean space. The dimensionality depends upon the character of the experiment. In many cases the points will be restricted to a plane and therefore embedding in two dimensions is required; in other cases embedding in three dimensions is appropriate. By a *finite relational structure* I mean a relational structure whose domain is finite. To give a simple example, suppose that A is the finite set of points and the judgments we have asked for are judgments of equidistance of points. Let E be the quaternary relation of equidistance. Then to say that the finite relational structure $\mathcal{U} = \langle A, E\rangle$ can be embedded in three-dimensional Euclidean space is to say that there exists a function φ defined on A such that φ maps A into the set of triples of real numbers and such that for every a, b, c, and d in A the following relation holds:

$$ab\ E\ cd \text{ iff } \textstyle\sum_{i=1}^{3}(\varphi_i(a) - \varphi_i(b))^2 = \sum_{i=1}^{3}(\varphi_i(c) - \varphi_i(d))^2,$$

where $\varphi_i(a)$ is the ith coordinate of $\varphi(a)$. Note that the mapping into triples of real numbers is just mapping visual points into a Cartesian representation of three-dimensional Euclidean space.

In principle, it is straightforward to answer the question raised by this embedding procedure. So that, given a set of data from an individual's visual judgments of equidistance between points, we can determine in a definite and constructive mathematical manner whether such embedding is possible.

Immediately, however, a problem arises. This problem can be grasped by considering the analogous physical situation. Suppose we are making observations of the stars and want to test a similar proposition, or some more complex proposition of celestial mechanics. We are faced with the problem recognized early in the history of astronomy, and also in the history of geodetic surveys, that the data are bound not to fit the theoretical model exactly. The classical way of putting this is that errors of measurement arise, and our problem is to determine if the model fits the data within the limits of the error of measurement. In examining data on the advancement of the perihelion of Mercury, which is one of the important tests of Einstein's general theory of relativity, the most tedious and difficult aspect of the data analysis is to determine whether the theory and the observations are in agreement within the estimated error of measurement.

Laplace, for example, used such methods with unparalleled success. He would examine data from some particular aspect of the solar system, for example, irregularities in the motion of Jupiter and Saturn, and would then raise the question of whether these observed irregularities were due to errors of measurement or to the existence of 'constant' causes. When the irregularities were too great to be accounted for by errors of measurement, he then searched for a constant cause to explain the deviations from the simpler model of the phenomena. In the case mentioned, the irregularities in the motion of Jupiter and Saturn, he was able to explain them as being due to the mutual gravitational attraction of the two planets. which had been ignored in the simple theory of their motion. But Laplace's situation is different from the present one in the following important respect. The data he was examining were already rendered in quantitative form and there was no question of having a numerical representation. Our problem is that we start from qualitative judgments and we are faced with the problem of simultaneously assigning a measurement and determining the error of that measurement. Because of the complexity and subtlety of the statistical questions concerning errors of measurement in the present setting, for purposes of simplification I shall ignore them, but it is absolutely essential to recognize that they must be dealt with in any detailed analysis of experimental data.

Returning to the formal problem of embedding qualitative relations among a finite set of points into a given space, it is surprising to find that the results of the kinds that are needed in the present context are not really present in the enormous mathematical literature on geometry. There is a very large literature on finite geometries; for example, Dembowski (1968) contains over 1200 references. Moreover, the tradition of considering finite geometries goes back at least to the beginning of this century. Construction of such geometries by Veblen and others was a fruitful source of models for proving independence of axioms, etc. On the other hand, the literature that culminates in Dembowski's magisterial survey consists almost entirely of projective and affine geometries that have a relatively weak structure. From a mathematical standpoint, such structures have been of considerable interest in connection with a variety of problems in abstract algebra. The corresponding theory of finite geometries of a stronger type, for example, finite Euclidean, finite elliptic, or finite hyperbolic geometries, is scarcely developed at all. As a result, the experimental literature does not deal directly with such finite geometries, although they are a natural extension of the weaker finite geometries on the one hand and finite measurement structures on the other.

A second basic methodological approach to the geometrical character of visual space is to assume that a standard metric representation already exists and then to examine which kind of space best fits the data. An excellent example of this methodology is to be found in various publications of Foley (1965, 1972). Foley shows experimentally that the size-distance invariance hypothesis, which asserts that the perceived size-distance ratio is equal to the physical size-distance ratio, is grossly incorrect At the same time he also shows that perceived visual angles are about ten percent greater than physical angles. These studies are conducted on the assumption that a variety of more primitive and elementary axioms are satisfied. In contrast, Luneburg (1948) assumes that the perceived visual angle equals the physical angle, that is, that the transformation between the two is conformal, but what is back of the use of this assumption is a whole variety of assumptions that both physical space and visual space are homogeneous spaces of constant curvature, that is, are Riemannian spaces, and essentially Luneburg does not propose to test in any serious way the many consequences implied by this very rich assumption of having a homogeneous space with constant curvature. In other words, in this second approach there is no serious attempt to provide tests that will show if all of the axioms that hold for a given type of space are satisfied.

A third approach is to go back to the well-known Helmholtz-Lie problem on the nature of space and to replace finiteness by questions of continuity and motion. In a famous lecture of 1854, Riemann (1866/1867) discussed the hypotheses on which the foundations of geometry lie. More than a decade later, Helmholtz (1868) responded in a paper entitled 'Über die Tatsachen, die der Geometrie zu Grunde liegen'. The basic argument of Helmholtz's paper was that, although arbitrary Riemannian spaces are conceivable, actual physical space has as an essential feature the free mobility of rigid bodies. From a mathematical standpoint, such motions are characterized in metric geometry as transformations of a space onto itself that preserve distances. Such transformations are called *isometries*. Because of the extensive mathematical development of the topic (for modern review, see Busemann, 1955, Section 48, or Freudenthal, 1965), an excellent body of formal results is available to use in the investigation of the character of visual space. Under various axiomatizations of the Helmholtz-Lie approach it can be proved that the only spaces satisfying the axioms are the following three kinds of elementary spaces: Euclidean, hyperbolic, and spherical.

From a philosophical standpoint, it is important to recognize that considerations of continuity and motion are probably more fundamental in the analysis of the nature of visual space than the mathematically more elementary properties of finite spaces. Unfortunately, I am not able to report any experimental literature that uses the Helmholtz-Lie approach as a way of investigating the nature of visual space, although it is implicit in some of the results reported below that it would be difficult to interpret the experimental results as satisfying an axiom of free mobility. Let me be clear on this point. Some of the experimental investigations lead to the result that visual space cannot be elementary in the sense just defined, but these investigations do not explicitly use the kind of approach to motion suggested by the rich mathematical developments that have followed in response to the Helmholtz-Lie problem.

A fourth approach that lies outside the main body of the literature to be considered in this paper is the recent approach through picture grammars and the analysis of perceptual scenes. Its growing literature has been in response especially to problems of pattern recognition that center on construction of computer programs and peripheral devices that have rudimentary perceptual capacities. Although this approach has a formal character quite different from the others considered and it has not been used to address directly the question about the Euclidean character of space, it should be mentioned because it does provide an approach that in many respects is very natural psychologically and that is in certain aspects more closely connected to the psychology of perception than most of the classical geometric approaches that have been used thus far in the analysis of visual space. (An elementary introduction and references

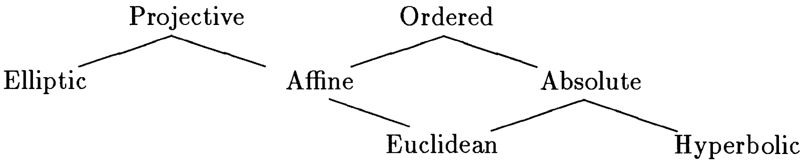

Figure 1. Hierarchy of geometries.

to the literature are to be found in Suppes and Rottmayer, 1974; an encyclopedic review is given by Fu, 1974.)

A typical picture grammar has the following character. Finite line segments or finite curves of a given length and with a given orientation are concatenated together as basic elements to form geometrical figures of greater complexity. A typical problem in the literature of pattern recognition is to provide such a concatenation (not necessarily one dimensional) so as to construct handwritten characters, or, as a specialized example that has received a fair amount of attention, to recognize handwritten mathematical symbols. These approaches are often labelled picture grammars because they adopt the approach used in mathematical linguistics for writing phrase-structure grammars to generate linguistic utterances. Picture grammars can in fact be characterized as context free, context sensitive, etc., depending upon the exact character of the rules of production. What is missing is the question, Can the set of figures generated by the picture grammars be embedded in Euclidean space or other metric spaces of an elementary character? This question would seem to have some conceptual interest from the standpoint of the theory of perception. It is clearly not of the same importance for the theory of pattern recognition. Picture grammars base perception on a set of primitive concepts that seem much more natural than the more abstract concepts familiar in classical geometry. They would seem to represent an excellent approach for exploration of the character of visual space but I am unable to cite references that test these ideas experimentally.

2. THE HIERARCHY OF GEOMETRIES

Those who have declared that visual space is not Euclidean have usually had a well-defined alternative in mind. The most popular candidates have been claims that visual space is either elliptic or hyperbolic, although some more radical theses are implicit in some of the experimental work.

How the various geometries are to be related hierarchically is not entirely a simple matter, for by different methods of specialization one may be obtained from another. A reasonably natural hierarchy for purposes of talking about visual space is shown in Figure 1. In the figure, I have also referred to geometries rather than to spaces, although from a certain conceptual standpoint the latter is preferable. I have held to the language of *geometries* in deference to tradition in the literature on visual space. The weakest geometry considered here is either projective geometry on the left-hand side at the top of the figure or ordered geometry at the right. There are various natural primitive concepts for projective geometry. Fundamental in any case is the concept of incidence and, once order is introduced, the concept of separation. In contrast, ordered geometry is based upon the single ternary relation of betweenness holding for three points in the fashion standard for Euclidean geometry, but of course axioms based only upon betweenness are weaker than those required for Euclidean geometry. Without entering into technical details, elliptic geometry of the plane is obtained from projective geometry by defining it as the geometry corresponding to the group of projective collineations that leave an imaginary ellipse invariant in the projective plane. Although elliptic geometry has been important in the consideration of visual space, as we shall see later, the details of elliptic geometry are complicated and subtle, and as far as I know have not actually been adequately studied in detail in relation to any serious body of experimental data.

Turning now to the right-hand side of Figure 1, affine geometry is obtained from ordered geometry by adding Euclid's axiom that, given a line and a point external to the line, there is at most one line (i) through the point, (ii) in the plane formed by the point and the line, and (iii) that does not meet the line. Going in the other direction from ordered geometry in Figure 1, we obtain absolute geometry by adding the concept of congruence of segments, which is just the notion of equidistance mentioned earlier. We add Euclid's axiom to absolute geometry to obtain Euclidean geometry and we add the negation of Euclid's axiom to absolute geometry to obtain hyperbolic geometry. These are the only two extensions of absolute geometry. Given the fundamental character of absolute geometry in relation to the claims often made that visual space is either Euclidean or hyperbolic, it is somewhat surprising that there has been no more detailed investigation experimentally of whether the axioms of absolute geometry hold for visual space.

There is another way of organizing the hierarchy of geometries in terms of metric spaces. Recall that a *metric space* is a pair $\langle A, d\rangle$ such that A is a nonempty set, d is a real-valued function defined on the Cartesian product $A \times A$, and for all a, b, c in A, Axiom 1. $d(a,a) = 0$ and if $a \neq b, d(a,b) > 0$;

Axiom 2. $d(a,b) = d(b,a)$;

Axiom 3. $d(a,b) + d(b,c) \geq d(\mathrm{a,c})$.

The elements of the set A are called *points*. The first axiom asserts that distances are positive, except for the distance between identical points, which is zero. The second axiom asserts that distance is symmetric; that is, it is a function only of the unordered pair of points, not a function of their order. The third axiom is the triangle inequality. Most of the metric spaces important for the theory of perception have the property that any two points can be joined by a segment. Such spaces are called metric spaces with additive segments. These spaces are naturally divided into two broad subclasses, affine metrics and coordinate-free metrics. By further specialization of each of these subclasses we are led naturally to the Euclidean, hyperbolic, and spherical spaces, as well as to generalizations of the Euclidean metric in terms of what are called Minkowski metrics. An important subclass of the coordinate-free metrics is the Riemannian metrics. It may be shown that the only spaces that are Riemannian and affine metric are either Euclidean or hyperbolic. We shall not use these concepts in detail, but it is to mention that this alternative hierarchy of metric spaces is as natural to use as the more classical hierarchy exhibited in Figure 1.

All of the concepts I have introduced in this brief survey of the hierarchy of geometries are familiar in the mathematical literature of geometry.

3. EXPERIMENTAL AND PHILOSOPHICAL ANSWERS

My main purpose in this section is to provide a survey of the answers that have been given. A summary is provided in Table 1.

The natural place to begin is with Euclid's *Optics*, the oldest extant treatise on mathematical optics. It is important to emphasize that Euclid's *Optics* is really a theory of vision and not a treatise on physical optics. A large number of the propositions are concerned with seeing from the standpoint of perspective in monocular vision. Indeed, Euclid's *Optics* could be characterized as a treatise on perspective within Euclidean geometry. The tone of Euclid's treatise can be seen from quoting the initial part, which consists of seven 'definitions'.

1. Let it be assumed that lines drawn directly from the eye pass through a space of great extent;

Name	Claim	Answer
Euclid (300 B.C.)	Theory of perspective	Yes
Reid (1764), Daniels (1972), Angell (1974)	Geometry of visibles is spherical	No
Blumenfeld (1913)	Parallel alleys not equal to equidistance alleys	No
Luneburg (1917, 1948, 1950)	Visual space is hyperbolic	No
Blank (1953, 1957, 1958a, 1958b, 1961)	Essentially the same as Luneburg	No
Hardy *et al.* (1953)	Essentially the same as Luneburg	No
Zajaczkowska (1956)	Positive results on experimental test of Luneburg theory	No
Schelling (1956)	Hyperbolic relative to given fixation point	No
Gogel (1956a, 1956b, 1963, 1964a, 1964b, 1965)	Equidistance tendency evidence for contextual geometry	No
Foley (1964, 1965, 1966, 1969, 1972)	Visual space is nonhomogeneous	No but
Indow (1967, 1968, 1974a, 1974b, 1975)	MDS methods yield good Euclidean fit	Not sure
Indow *et al.* (1962a, 1962b, 1963)	Close to Indow	Not sure
Nishikawa (1967)	Close to Indow	Not sure
Matsushima and Noguchi (1967)	Close to Indow	Not sure
Grünbaum (1963)	Questions the theory of Luneburg	Yes
Strawson (1966)	Phenomenal geometry is Euclidean	Yes

Table 26.1. Is Visual Space Euclidean?

2. and that the form of the space included within our vision is a cone, with its apex in the eye and its base at the limits of our vision;
3. and that those things upon which the vision falls are seen, and that those things upon which the vision does not fall are not seen;
4. and that those things seen within a larger angle appear larger, and those seen within a smaller angle appear smaller, and those seen within equal angles appear to be of the same size;
5. and that those things seen within the higher visual range appear higher, while those within the lower range appear lower;
6. and, similarly, that those seen within the visual range on the right appear on the right, while those within that on the left appear on the left
7. but that things seen within several angles appear to be more clear.

(The translation is taken from that given by Burton in 1945.)

The development of Euclid's *Optics* is mathematical in character, but it is not axiomatic in the same way that the *Elements* are. For example, later Euclid proves two propositions, 'to know how great is a given elevation when the sun is shining' and 'to know how great is a given elevation when the sun is not shining'. As would be expected, there is no serious introduction of the concept of the sun or of shining but they are treated in an informal, commonsense, physical way with the essential thing for the proof being rays from the sun falling upon the end of a line. Visual space is of course treated by Euclid as Euclidean in character.

The restriction to monocular vision is one that we shall meet repeatedly in this survey. However, it should be noted that Euclid proves several propositions involving more than one eye; for example, 'If the distance between the eyes is greater than the diameter of the sphere, more than the hemispheres will be seen'. Euclid is not restricted to some simple geometric optics but is indeed concerned with the theory of vision, as is evident from the proposition that 'if an arc of a circle is placed on the same plane as the eye, the arc appears to be a straight line'. This kind of proposition is a precursor of later theories—for example, that of Thomas Reid—which emphasize the non-Euclidean character of visual space.

I skip rapidly through the period after Euclid to the eighteenth century, not because there are not matters of interest in this long intervening period but because there do not seem to be salient changes of opinion about the character of visual space, or at least if there are they are not known to me. I looked, for example, at the recent translation by David C. Lindberg (1970) of the thirteenth-century treatise *Perspectiva Communis* of John Pecham and found nothing to report in the present context, although the treatise itself and Lindberg's comments on it are full of interesting matter of great importance concerning other questions in optics, as, for example, theories about the causes of light.

Newton's *Opticks* (1704/1931) is in marked contrast to Euclid's. The initial definitions do not make any mention of the eye until Axiom VIII, and then in very restrained fashion. Almost without exception, the propositions of Newton's optics are concerned with geometrical and especially physical properties of light. Only really in several of the Queries at the end are there any conjectures about the mechanisms of the eye, and these conjectures do not bear on the topic at hand.

Five years after the publication of the first edition of Newton's *Opticks*, Berkeley's *An Essay Towards a New Theory of Vision* (1709/1901) appeared in 1709. Berkeley does not really have much of interest to say about the geometry of visual space, except in a negative way. He makes the point that distance cannot be seen directly and, in fact, seems to categorize the perception of distance as a matter of tactile rather than visual sensation because the muscular convergence of the eyes is tactile in character. He emphatically makes the point that we are not able geometrically to observe or compute the optical angle generated by a remote point as a vertex with sides pointing toward the centers of the two eyes. Here is what he says about the perception of optical angles. "Since therefore those angles and lines are not themselves perceived by sight, it follows,... that the mind does not by them judge the distance of objects" (# 13). What he says about distance he also says about magnitude not being directly perceived visually. In this passage (# 53), he is especially negative about trying to use the geometry of the visual world as a basis for visual perception.

It is clear from these and other passages that for Berkeley visual space is not Euclidean because there is no proper perception of distance or magnitude; at least, visual space is not a three-dimensional Euclidean space. What he seems to say is sufficiently ambiguous as to whether one should argue that it is at least a two-dimensional Euclidean space. My own inclination is to judge that his views on this are more negative than positive. Perhaps a sound negative argument can be made up from his insistence on there being a minimum visible. As he puts it, "It is certain sensible extension is not infinitely divisible. There is a minimum tangible, and a minimum visible, beyond which sense cannot perceive. This everyone's experience will inform him" (# 54).

In fact, toward the end of the essay, Berkeley makes it clear that even two-dimensional geometry is not a proper part of visual space or, as we might say, the visual field. As he says in the final paragraph of the essay, "By this time, I suppose, it is clear that neither abstract nor visible extension makes the object of geometry."

Of much greater interest here is Thomas Reid's *Inquiry into the Human Mind*, first published in 1764 (1764/1967). Chapter 6 deals with seeing, and Section 9 is the celebrated one entitled 'Of the geometry of visibles'. It is sometimes said that this section is a proper precursor of non-Euclidean geometry, but if so, it must be regarded as an implicit precursor because the geometry explicitly discussed by Reid as the geometry of visibles is wholly formulated in terms of spherical geometry, which had of course been recognized as a proper part of geometry since ancient times. The viewpoint of Reid's development is clearly set forth at the beginning of the section: "Supposing the eye placed in the centre of a sphere, every great circle of the sphere will have the same appearance to the eye as if it was a straight line; for the curvature of the circle being turned directly toward the eye, is not perceived by it. And, for the same reason, any line which is drawn in the plane of a great circle of the sphere, whether it be in reality straight or curve, will appear to the eye." It is important to note that Reid's geometry of visibles is a geometry of monocular vision. He mentions in other places binocular vision, but the detailed geometrical development is restricted to the geometry of a single eye. The important contrast between Berkeley and Reid is that Reid develops in some detail the geometry in a straightforward, informal, mathematical fashion. No such comparable development occurs in Berkeley.

Daniels (1972) has argued vigorously that Reid's geometry of visibles is not simply a use of spherical geometry but is an introduction by Reid of a double elliptic space. A similar argument is made by Angell (1974). I am sympathetic with these arguments, but it seems to me that they go too far and for a fairly straightforward reason not discussed by either Daniels or Angell. Let us recall how elliptic geometry was created by Felix Klein at the end of the nineteenth century. He recognized that a natural geometry, very similar to Euclidean geometry or hyperbolic geometry could be obtained from spherical geometry by identifying antipodal points as a single point. The difficulty with spherical geometry as a geometry having a development closely parallel to that of Euclidean geometry is that two great circles, which correspond to lines, have two points, not one point, of intersection. However, by identifying the two antipodal points as a single point, a fair number of standard Euclidean postulates remain valid. It is quite clear that no such identification of antipodal points was made by Reid, for he says quite clearly in the fifth of his propositions, 'Any two right lines being produced will meet in two points, and mutually bisect each other'. This property of meeting in two points rather than one is what keeps his geometry of visibles from being a proper elliptic geometry and forces us to continue to think of it in terms of the spherical model used directly by Reid himself.

In spite of the extensive empirical and theoretical work of Helmholtz on vision, he does not have a great deal to say that directly bears on this question, and I move along to experiments and relevant psychological theory in the twentieth century. The first stopping point is Blumenfeld (1913).

Blumenfeld was among the first to perform a specific experiment to show that, in one sense, phenomenological visual judgments do not satisfy all Euclidean properties. Blumenfeld performed experiments with so-called parallel and equidistance alleys. In a darkened room the subject sits at a table, looking straight ahead, and he is asked to adjust two rows of point sources of light placed on either side of the normal plane, i.e., the vertical plane that bisects the horizontal segment joining the centers of the two eyes. The two furthest lights are fixed and are placed symmetrically and equidistant from the normal plane. The subject is then asked to arrange the other lights so that they form a parallel alley extending toward him from the fixed lights. His task is to arrange the lights so that he perceives them as being straight and parallel to each other in his visual space. This is the task for construction of a parallel alley. The second task is to construct a distance alley. In this case, all the lights except the two fixed lights are turned off and a pair of lights is presented, which are adjusted as being at the same physical distance apart as the fixed lights—the kind of equidistance judgments discussed earlier. That pair of lights is then turned off and another pair of lights closer to him is presented for adjustment, and so forth. The physical configurations do not coincide, but in Euclidean geometry straight lines are parallel if and only if they are equidistant from each other along any mutual perpendiculars. The discrepancies observed in Blumenfeld's experiment are taken to be evidence that visual space is not Euclidean. In both the parallel-alley and equidistance-alley judgments the lines diverge as you move away from the subject, but the angle of divergence tends to be greater in the case of parallel than in the case of equidistance alleys. The divergence of the alleys as one moves away from the subject has been taken by Luneburg to support his hypothesis that visual space is hyperbolic.

In fact, Luneburg, in several publications in the late forties, has been by far the strongest supporter of the view that visual space is hyperbolic. He, in conjunction with his collaborators, has set forth a detailed mathematical theory of binocular vision and at the same time has generated a series of experimental investigations to test the basic tenants of the theory.

In many respects, Luneburg's article (1947) remains the best detailed mathematical treatment of the theory of binocular vision. Without extensive discussion, Luneburg restricts himself to Riemannian geometries of constant curvature in order to preserve rigid motions, that is, free mobility of rigid bodies. Luneburg develops in a coordinate system natural for binocular vision the theory of Riemannian spaces of constant curvature in a quite satisfactory form, although an explicit axiomatic treatment is missing. On the other hand, he nowhere examines with any care or explicitness the more general and primitive assumptions that lead to assuming that visual space is a Riemannian space of constant curvature. After these general developments he turns to the detailed arguments for the view that the appropriate space of constant curvature for visual space is hyperbolic. It is not possible to enter into the details of Luneburg's argument here, but he bases it on three main lines of considerations, all of which have had a great deal of attention in the theory of vision: first, the data arising from the frontal-plane horopter where curves which appear as straight are physically curved (data on these phenomena go back to the time before Helmholtz); second, the kind of alley phenomena concerning judgments of parallelness mentioned earlier; and, third, accounting for judgments of distorted rooms in which appropriate perspective lines are drawn and which consequently appear as rectangular or regular (here, Luneburg draws on some classic and spectacular demonstrations by A. Ames, Jr.). One of the difficulties of this field is that the kind of detailed mathematical and quantitative arguments presented by Luneburg in connection with these three typical kinds of problems are not satisfactorily analyzed in the later literature. Rather, new data of a different sort are presented to show that different phenomena argue against Luneburg's hypothesis that visual space is hyperbolic.

Luneburg died in 1949, but a number of his former students and collaborators have continued his work and provided additional experimental support as well as additional mathematically based arguments in favor of his views. I refer especially to Blank (1953, 1957, 1958a, 1958b, 1961) and Hardy, Rand, Rittler, Blank, and Boeder (1953), although this is by no means an exhaustive list. Another positive experimental test was provided by Zajaczkowska (1956).

Schelling (1956) agrees with Luneburg but makes an important point of modification, namely, the metrics of negative curvature—that is, of the hyperbolic spaces that Luneburg argues for—are essentially momentary metrics. At a given instant the eye has a certain fixation point, and relative to this fixation point Luneburg's theory is, according to Schelling. probably approximately correct, but the applicability of the theory is severely restricted because the eyes are normally moving about continuously and the points of fixation are continually changing. This fundamental fact of change must he taken account of in any fully adequate theory.

Gogel (1956a, 1956b, 1963, 1964a, 1964b, 1965) has studied what is called the equidistance tendency, or what in the context of this paper we might term the Berkeley tendency. Remember that Berkeley held that distance was not a visual idea at all but derived from the tactile sense. Without entering into a precise analysis of Berkeley's views, Gogel has provided an important body of evidence that when other cues are missing there is a strong tendency to view objects as being at the same distance from the observer. These careful and meticulous studies of Gogel are important for establishing not only the equidistance tendency but also its sensitivity to individual variation, on the one hand, and to the presence of additional visual cues on the other. The equidistance tendency is certainly present as a central effect. but any detailed theory of visual space has a bewildering complexity of contextual and individual differences to account for, and it seems to me that Gogel's experiments are essentially decisive on this point. In the papers referred to, Gogel does not really give a sharp answer to the question about the character of visual space, but I have listed him in Table 1 because it seems to me that the impact of his studies is to argue strongly for skepticism about fixing the geometry of visual space very far up in the standard hierarchy and, rather, to insist on the point that the full geometry is strongly contextual in character and therefore quite deviant from the classical hierarchy.

A number of interesting experimental studies of the geometry of visual space have been conducted by John Foley. In Foley (1964) an experiment using finite configurations of small point sources of light was conducted to test the Desarguesian property of visual space. (Of course, the property was tested on the assumption that a number of other axioms were valid for visual space.) The results confirmed the Desarguesian property for most observers but not for all. In Foley (1966), perceived equidistance was studied as a function of viewing distance. Like most of Foley's experiments, this was conducted in the horizontal eye-level plane. The locus of perceived equidistance was determined at distances of 1.2, 2.2, 3.2, and 4.2 meters from the observer. As in other Foley experiments, the stimuli were small, point-like light sources viewed in complete darkness. The observer's head was held fixed but his eyes were permitted to move freely. There were five lights, one in the normal plane, which was fixed, and two variable lights on each side of the normal plane at angles of 12 degrees and 24 degrees with respect to the normal plane. The locus of perceived equidistance was found to be concave toward the observer at all distances. Perhaps most importantly the locus was found to vary with viewing distance, which indicates that the visual space does not depend on the spatial distribution of retinal stimulation alone. Again, there is here a direct argument for a contextual geometry and results are not consistent with Luneburg's theory. The equidistance judgments were of the following sort: A subject was instructed to set each of the lights, except the fixed light, in the normal plane to be at the same distance from himself as the fixed light. Thus, it should appear to him that the lights lie on a circle, with himself as observer at the center. The important point is that for none of the ten subjects in the experiment did the judgments of the locus for equidistance lie on the Vieth-Mueller horopter or circle mentioned earlier as one of the supporting arguments for Luneburg's theory. Also important for the fundamental geometry of visual space is the fact that the loci determined by the observers were not symmetric about the normal plane.

Foley's (1972) study shows experimentally that, on the one hand, the size-distance invariance hypothesis is incorrect, and that in fact the ratio of perceived frontal extent to perceived egocentric distance greatly exceeds the physical ratio, while, on the other hand, perceived visual angles are quite close to physical ones. These results, together with other standard assumptions, are inconsistent with the Luneburg theory that visual space is hyperbolic. Foley describes the third experiment in this paper in the following way:

> How can it be that in the primary visual space reports of perceived size-distance ratio are not related to reports of perceived visual angle in a Euclidean way? One possibility is that the two kinds of judgments are in part the product of different and independent perceptual processes... . The results are consistent with the hypothesis that the two kinds of judgments are the product of independent processes. They also show that no one geometrical model can be appropriate to all stimulus situations, and they suggest that the geometry may approach Euclidean geometry with the introduction of cues to distance.

Again, there is in Foley's detailed analysis a strong case for a contextual geometry. A number of other detailed experimental studies of Foley that have not been referenced here build a case for the same general contextual view, which I discuss in more detail below.

A number of detailed investigations on the geometry of visual space have been conducted by Tarow Indow (1967, 1968, 1974a, 1974b, 1975) and other Japanese investigators closely associated with him (Indow et al., 1962a, 1962b, 1963; Matsushima and Noguchi, 1967; Nishikawa, 1967). They have found, for example, that multidimensional scaling methods

(MDS), which have been intensively developed in psychology over the past decade and a half, in many cases yield extremely good fits to Euclidean space. Indow has experimentally tested the Luneburg theory based upon the kind of alley experiments that go back to Blumenfeld (1913). As might be expected, he duplicates the result that the equidistance alleys always lie outside the parallel alleys, which under the other assumptions that are standard implies that the curvature of the space is negative and therefore it must be hyperbolic. But Indow (1974a,b) properly challenges the simplicity of the Luneburg assumptions, especially the constancy of curvature. It is in this context that he has also tried the alternative approach of determining how well multidimensional scaling will work to fit a Euclidean metric. As he emphasizes also, the Luneburg approach is fundamentally based upon differential geometry as a method of characterizing Riemannian spaces with constant curvature, but for visual judgments it is probably more appropriate to depend upon the judgments in the large and therefore upon a different conceptual basis for visual geometry. Throughout his writings, Indow recognizes the complexity and difficulty of reaching for any simple answer to give the proper characterization of visual space. The wealth of detail in his articles and those of his collaborators is commended to the reader who wants to pursue these matters in greater depth.

In his important book on the philosophy of space and time, Grünbaum (1963) rejects the Luneburg theory and affirms that, in order to yield the right kinds of perceptual judgments, visual space must be Euclidean. His argument is rather brief and I shall not examine it in any detail. It would be my own view that he has not given proper weight to the detailed experimental studies or to the details of the various theoretical proposals that have been made.

I close this survey by returning to a philosophical response to the question, that of Strawson (1966) in his book on Kant's *Critique of Pure Reason*. From the standpoint of the large psychological literature I have surveyed, it is astounding to find Strawson asserting as a necessary proposition that phenomenal geometry is Euclidean. The following quotation states the matter bluntly:

> With certain reservations and qualifications, to be considered later, it seems that Euclidean geometry may also be interpreted as a body of unfalsifiable propositions about phenomenal straight lines, triangles, circles, etc.; as a body of a priori propositions about spatial appearances of these kinds and hence, of course, as a theory whose application is restricted to such appearances. (p. 286)

The astounding feature of Strawson's view is the absence of any consideration that phenomenal geometry could be other than Euclidean and that it surely must be a matter, one way or another, of empirical investigation to determine what is the case. The qualifications he gives later do not bear on this matter but pertain rather to questions of idealization and of the nature of constructions, etc. The absence of any attempt to deal in any fashion whatsoever with the large theoretical and experimental literature on the nature of visual space is hard to understand.

4. SOME REMARKS ON THE ISSUES

In this final section, I center my remarks around three clusters of issues. The first is concerned with the contextual character of visual geometry, the second with problems of distance perception and motion, and the third with the problem of characterizing the nature of the objects of visual space.

Contextual geometry. A wide variety of experiments and ordinary experience as well testify to the highly contextual character of visual space. The presence or absence of 'extraneous' points can sharply affect perceptual judgments. The whole range of visual illusions, which I have not discussed here, provides a broad body of evidence for the surprising strength of these contextual effects.

As far as I can tell, no one has tried seriously to take account of these contextual effects from the standpoint of the axiomatic foundations of visual geometry. In a way it is not surprising, for the implications for the axiomatic foundations are, from the ordinary standpoint, horrendous. Let us take a simple example to illustrate the point.

In ordinary Euclidean geometry, three points form an isosceles triangle just when two sides of the triangle are of the same length. Suppose now that Euclidean geometry had the much more complicated aspect that whether a triangle were isosceles or not depended not simply on the configuration of the three points but also on whether there was a distinguished point lying just outside the triangle alongside one of the dual sides. This asymmetry may well make the visual triangle no longer isosceles. This is but one simple instance of a combinatorial nightmare of contextual effects that can easily be imagined and, without much imagination or experimental skill, verified as being real.

What are we to say about such effects? It seems to me the most important thing is to recognize that perceptual geometry is not really the same as classical geometry at all, but in terms of the kinds of judgments we are making it is much closer to physics. Consider, for example, the corresponding situation with bodies that attract each other by gravitation. The introduction of a third body makes all the difference to the motions of the two original bodies and it would be considered bizarre for the situation to be otherwise. This also applies to electromagnetic forces, mechanical forces of impact, etc. Contextual effects are the order of the day in physics, and the relevant physical theories are built to take account of such effects.

Note that physical theories depend upon distinguished objects located in particular places in space and time. Space-time itself is a continuum of undistinguished points, and it is characteristic of the axiomatic foundations of classical geometry that there are no distinguished points in the space. But it is just a feature of perception that we are always dealing with distinguished points which are analogous to physical objects, not geometrical points. Given this viewpoint, we are as free to say that we have contextual effects in visual geometry as we are to make a similar claim in general relativity due to the presence of large masses in a given region.

Interestingly enough, there is some evidence that as we increase the visual cues, that is, we fill up the visual field with an increasingly complex context of visual imagery, the visual space becomes more and more Euclidean. It is possible that we have here the exact opposite of the situation that exists in general relativity. In the case of perception it may be that spaces consisting of a very small number of visible points may be easily made to deviate from any standard geometry.

The geometric viewpoint can be brought into close alignment with the physical one, when the embedding of finite sets of points in some standard geometry is taken as the appropriate analysis of the nature of visual space. This approach was mentioned earlier and is implicit in some of the experimental literature discussed. It has not sufficiently been brought to the surface, and the full range of qualitative axioms that must be satisfied for the embedding of a finite collection of points in a unique way in a given space, whether Euclidean, hyperbolic, elliptic, or what not, needs more explicit and detailed attention.

It also seems satisfactory to avoid the problems of contextual effects in initial study of this kind by deliberately introducing symmetries and also certain special additional assumptions such as quite special relations of a fixed kind to the observer. The many different experimental studies and the kind of mathematical analysis that has arisen out of the Luneburg tradition suggest that a good many positive and almost definitive results could be achieved under special restrictive assumptions. It seems to me that making these results as definitive as possible, admitting at the same time their specialized character and accepting the fact that the general situation is contextual in character, is an appropriate research strategy.

It also seems to me likely that for these special situations one can give a definitely negative answer to the question, Is visual space Euclidean?, and respond that, to high approximations, in many special situations it is hyperbolic and possibly in certain others elliptic in character. This restricted answer is certainly negative. A general answer at the present time does not seem available as to how to characterize the geometry in a fully satisfactory way that takes account of the contextual effects that are characteristic of visual illusions, equidistance tendencies, etc.

Distance perception and motion. As indicated earlier in the brief discussion of the Helmholtz-Lie problem, most of the work surveyed in the preceding section has not taken sufficient account of problems of motion. There is an excellent survey article of Foley (1978) on distance perception which indicates that eye motion during the initial stage of focusing on an object is especially critical in obtaining information about perceptual distance. In spite of the views of Berkeley, philosophical traditions in perception have tended to ignore the complicated problems of motion of the eyes or head as an integral part of visual perception, but the most elementary considerations are sufficient to demonstrate their fundamental importance. It was a fundamental insight of Luneburg to recognize that it is important to characterize invariance properties of motions of the eyes and head that compensate each other. The deeper aspects of scanning as determining the character of the visual field have not really been studied in a thoroughly mathematical and quantitative fashion, and there is little doubt in my mind that this is the area most important for future developments in the theory of visual perception. We should, I would assume, end up with a kinematics of visual perception replacing the geometry of visual perception. For example, Lamb (1919) proves that under Donders' law, which asserts that the position of the eyeball is completely determined by the primary position and the visual axis aligned to the fixation point, it is not possible for every physically straight line segment to be seen as straight. This kinematical theorem of Lamb's, which is set forth in detail in Roberts and Suppes (1967), provides a strong kinematical argument against the Euclidean character of visual space. I cite it here simply as an example of the kind of results that one should expect to obtain in a more thoroughly developed kinematics of visual perception.

Objects of visual space. Throughout the analysis given in this paper the exact characterization of what are to be considered as the objects of visual space has not been settled in any precise or definitive way. This ambiguity has been deliberate because the wide range of literature to which I have referred does not have a settled account of what are to be regarded as the objects of visual space. The range of views is extreme—from Berkeley, who scarcely even wants to admit a geometry of pure visual space, to those who hold that visual space is simply a standard Euclidean space and there is little real distinction between visual objects and physical objects. In building up the subject axiomatically and systematically, clearly some commitments are needed, and yet it seems that one can have an intelligible discussion of the range of literature considered here without having to fix upon a precise characterization, because there is broad agreement on the look of things in the field of vision. From the standpoint of the geometry of visual space, we can even permit such wide disagreement as to whether the objects are two dimensional or three dimensional in order to discuss the character of the geometry. Thomas Reid would lean strongly toward the two-dimensional character of visual space. Foley would hold that visual space is three dimensional; note, however, that most of his experiments have been restricted to two dimensions. At the very least, under several different natural characterizations of the objects of visual space it is apparent that strong claims can be made that visual space is not Euclidean, and this is a conclusion of some philosophical interest.

27

DAVIDSON'S VIEWS ON PSYCHOLOGY AS A SCIENCE

More than two decades ago Davidson and I together conducted several experiments on decision making. We have not talked much about psychology for many years and, as will be apparent from my remarks in this short article, our views about psychology as a science and, indeed, our views about science in general, diverge. All the same, I find what Davidson has to say about psychology enormously interesting and stimulating. I have confined my comments to three articles, 'The Material Mind' (*MM*), 'Thought and Talk' (*TT*), and 'Psychology as Philosophy' (*PP*).[1] Certainly other articles of Davidson's are relevant to the themes advanced in these three, but the limitation is not unreasonable for the restricted purposes of my analysis.

I regard *MM* as a classic that should be required reading for a variety of folk, and I strongly agree with its final sentence: 'There is no important sense in which psychology can be reduced to the physical sciences.'

On the other hand, I do not agree with many of Davidson's more detailed arguments and conclusions. I have selected five theses that I

*Reprinted from *Essays on Davidson: Actions and events* (ed. by V. Vermazen and M.B. Hintikka), 1985, pp. 183-194. Oxford, England: Clarendon Press. I have benefited from a variety of critical remarks on an earlier draft by Philip Staines and my wife, Christine.

[1] Reprinted in Davidson (1980), pp. 245–259, Davidson (1984), pp. 155–170, and Davidson (1980), pp. 229–239.

think are defensible and that more or less contradict views that Davidson has advanced in one place or another in one of the three articles, including the printed discussion of *PP*.

Before turning to the first thesis, there is a general methodological point I want to make about the three articles and my comments. Nothing is proved in detail. The arguments are not complete. In arguing that psychological concepts are not connected in a lawlike way with physical concepts, Davidson likes to take as a parallel example the semantic impossibility of giving within a fairly rich language a definition of truth for that language (*MM*, pp. 249-50). But there is an important methodological difference about this example. Circa 1980 it can be confidently described in a few sentences because the underlying semantic theory was given such a satisfactory and explicit form much earlier by Tarski and others. Davidson's arguments (or mine) are not cast out of the same mold, and I miss in his arguments and analysis the formulation of problems and issues he cannot solve. It is hard to believe he regards his arguments as definitive for fixing, as Kant might put it, the possible limits of any future psychology. On the other hand, he gives few if any hints about how he thinks the arguments can be made more formal and explicit. I hope that he and I may agree that this is work yet to be done.

I turn now to the five theses.

(1) *It is common in physics as well as in psychology to study systems that are not closed, that are not deterministic, and that are holistic in character.*

I have mentioned in this first thesis three properties that Davidson (*PP*, pp. 229-30) gives to discriminate physics from psychology. A casual perusal of the *Physical Review* would substantiate the claim that determinism is as dead in physics as it is in psychology. (Even Einstein never really seemed to believe in determinism.) The present strong interest in astrophysics—some physicists regard it as the most promising current area of research—and the widespread renewed interest in theories of space provide evidence enough that physicists do not deal primarily with closed systems. A concern with holistic theory is found in current views on the beginnings of the universe or, to take a less exotic example, in the great emphasis on field theories in physics for the past quarter of a century.[2]

[2] Following is a quotation from a currently fashionable book (Hawking & Ellis (1973)): 'In fact it may not be possible to isolate a system from external matter fields. Thus for example in the Brans-Dicke theory there is a scalar field which is non-zero everywhere' (p. 64). Here is C. Truesdell (1968) on the rather small importance of experiments for the development of rational mechanics. This passage says

In *MM* (p. 245), Davidson sets aside the indeterminism of quantum mechanics (but not of astrophysics) as part of a fanciful tale about complete physical specification of a person, but in *PP* (p. 231) he declares as irrelevant the possibly irreducible probabilistic character of microphysics. In another passage (*MM*, p. 246), he says that the assumption of determinism for macrophysics is not essential to his argument. I hope that in the future he will elaborate on this point, for it seems to me that from today's perspective it is only potted physics, of the sort taught undergraduates, that is deterministic. There is even a substantial literature on the indeterminism of classical mechanics.[3] From a purely psychological standpoint, microphysics does seem relevant because of the enormous sensitivity of the visual and olfactory senses to essentially a quantum of light in the one case and a few molecules in the other. These two examples are easily multiplied because of the near-molecular level of many physiological phenomena that obviously interact with psychological states.

For these and other reasons I shall in the sequel discuss quantum mechanics as a relevant physical theory.

Perhaps the best argument against closed systems in physics is the prominent place in quantum mechanics given to disturbances of a system due to measurement. To put the case in most extreme form, it might be said that there may well be closed systems in physics but we shall never be able to observe them. Moreover, parallel to what Davidson says about psychological phenomena (*PP*, p. 230), quantum phenomena are observed in terms of macroscopic concepts that are foreign to microconcepts. To use G. E. Moore's concept of supervenience (as Davidson does in *MM*, pp. 253-4), we might argue for the supervenience, from a human standpoint, of the microphysical with respect to the macrophysical. One of the benefits

some of the things about mechanics that Davidson says about psychology. I cite it as part of my general argument that Davidson tends to separate physics and psychology methodologically and theoretically more than I think is warranted. 'Without *experience* there would be no rational mechanics, but I should mislead you if I claimed that experiment, either now or 200 years ago, had greatly influenced those who study rational mechanics. In this connection experiment, like alcohol, is a stimulant to be taken with caution. To consult the oracle of a fine vintage at decent intervals exhilarates, but excess of the common stock brings stupor. Students of rational mechanics spend much effort thinking *how materials might possibly behave.* These thoughts have not been fruitless of information on how some materials do behave. Real materials are not naive; neither are they irrational' (p. 357).

[3] For a beautiful instance, see Gale (1952). King Oscar of Sweden's prize was given to Henri Poincaré for his work on the n-body problem in classical mechanics, in so far as it solved Laplace's problem of the stability of the solar system. What Poincaré's and subsequent results showed is that the necessarily infinite series methods of solution prevent us from having a satisfactory deterministic solution of even the three-body problem.

of having a more formal version of Davidson's arguments on these matters would be being able to examine the extent to which a structurally similar argument could be made for the relation between quantum mechanics and classical physics.

It is a favorite theme of mine—I do not have time to expand upon it here—that physics is becoming like psychology. In this sense, some of the pessimism Davidson expresses about psychology I would extend to contemporary physical theory, but this is not the point he wants to make. His grounds for differentiation of physics and psychology in terms of closed systems, determinism, and holistic properties are, I believe, hard to make a case for in detail. In his wide-ranging criticism of the possibility of fundamental psychological theory, or at least fundamental theory about propositional attitudes, Davidson has thrown out the physical baby with the psychological bath water. He seems to want to impose a standard for fundamental scientific theory that is satisfied neither by physics nor by psychology.

(2) *Much of modern physical theory is intensional in expression and the reports of physical experiments are intensional accounts of human activity that cannot properly be expressed in extensional form.*

Thus, once again I accept much of Davidson's thesis about psychology, but it is not a thesis that strongly differentiates psychology from physics. I certainly grant that the concepts that are intensional in psychology are often different at the theoretical level from those in physics. Thus there is no natural place in physical theory for concepts of purpose and desire, but there is a natural place for a concept closely related to belief, that of probability, and if we adopt a thoroughly subjective view toward probability, the same concept would apply to belief that applies to the expression of probability in physics. The important point, however, is that the use of probability concepts in physics is essential to almost all modern theory and is at the same time, thoroughly intensional in character. Many familiar examples show that probability statements create intensional contexts. We may calculate in a given theory that the probability of two events A and B being identical is some number between 0 and 1/2, but without calculation the probability that A is identical to A is 1.[4] Moreover, so distinguished a physicist as Eugene Wigner traces the problem of measurement in quantum mechanics all the way back to the (intentional) consciousness of the observer.

Perhaps the more important point is that in the standard accounts of physical experiments the use of intensional language is widespread and, in

[4] I have expanded upon this point in Suppes (1974d).

my view, uneliminable. Philosophers of science have generally neglected the details of actual experiments or the language in which experiments are reported. Let me give a couple of examples of such intensionality.

Here is Henri Becquerel in 1896 (1964 translation):

> I then attempted to transmit a new activity to these substances by various known procedures. I heated them in the presence of the photographic plate without heating the latter, and I obtained no impression (p. 17).

Becquerel is perhaps especially apposite to quote because his classic experiments on establishing the existence of radioactivity constituted a major step in building the current edifice that has destroyed the classical deterministic view of physics.

Here is Ernest Rutherford in 1900 using the plain man's concept of expectation (reprinted in 1964): 'If the radiation is of one kind, we should expect the rate of discharge (which is proportional to the intensity of the radiation) to diminish in geometrical progression with the addition of equal thicknesses of paper. The above figures show this is approximately the case' (p. 27).

I can see little difference between the theoretical status of trying to infer something about the probabilistic structure of beliefs of an individual and the probabilistic structure of decay in radioactive atoms. Neither structure is amenable to direct observation; both require elaborate and subtle experimental procedures of an intentional kind to test significant aspects of theoretical claims.

(3) *Animals have beliefs.*

In *TT* Davidson gives several different arguments why dogs and other mammals that do not talk cannot have beliefs. On page 170, he succinctly summarizes his main points: (1) The idea of belief comes from the interpretation or understanding of language; (2) a creature that has beliefs must have the concept of belief; (3) a creature that has beliefs must also have the concept of error of belief and thus the concepts of truth and falsity. I find these arguments unpersuasive and I shall try to say why. There is, however, a more general issue I want to comment on first. Certainly most plain men believe that dogs, monkeys, and other primates have beliefs and are capable of thinking about a certain range of problems. As some philosophers in the recent past might put it, it is analytic that animals have beliefs because of this widespread common opinion and common acceptance of the 'fact' in casual conversation and the like. I certainly do not oppose going against the grain of the plain man when scientific theory demands it. There are plenty of examples of importance to be cited that require it. But to go against the grain requires a detailed theory with an articulation of concepts in a systematic structure. This, it seems to me, Davidson has not provided.

A dog waits at the door. We say that he expects his mistress to arrive, or we may say that he believes that his mistress will arrive soon. A cat meows at the door. We say that he thinks it is time to be fed. The monkey grabs a stick in order to reach a banana outside the cage. We may say that he grabs the stick and uses it because he expects to be able to reach the banana, or, put another way, he believes that he can reach the banana. It seems to me that we can stipulate, in order to agree with Davidson (not that I do), that the concept of belief arose in connection with the interpretation of language, but that does not mean at all that its use is now restricted to a linguistic context. We could, on the same principles, say that there can be no proper non-human physical concept of force, because we can maintain with Jammer (1957) and others that the initial primitive concept of force is that of muscular force. There have been occasional attempts in the history of physics to exclude the concept of force and to reduce mechanics to pure kinematics, but these attempts at elimination seem to me as unsuccessful as those aimed at a similar elimination of the concept of belief for animals.

It simply is the case that people talk about beliefs, thoughts, and expectations of animals in the style of my simple examples, and it seems to me there is a natural and straightforward interpretation of these uses that places them outside the restrictive framework that Davidson would like to impose on the concept of belief.

Let me now try to deal more directly with Davidson's main points cited above. The analysis just given, favored by animal-lovers everywhere, he may set aside as being mistaken and in need of a fundamental revision, for which he has written the prolegomena. That his own views require revision, in order to be viable as a relevant theory, seems to me to be most directly seen by considering an array of data from developmental psychology, including those on language acquisition. A variety of data shows indisputably that only gradually does a child master either language comprehension or language production, but his intentional motor behavior is well developed much sooner. I would say that as the child learns to crawl about, he early develops beliefs concerning what is and is not feasible, what can be ventured and what not. If we turn to his language productions of single-word utterances around 22 months, it is difficult to hold that at this stage his beliefs have the properties Davidson alleges are necessary for belief. It is even difficult for me to believe that these properties are there when he is 36 months and babbles away in two-, three-, and four-word utterances. What general concept of belief does he have? What concept of truth? On the holistic theory of language, meaning, and interpretation advanced in *MM* (pp. 256-7), it is not easy to see how a child could acquire beliefs at all. Short of his giving us the details of an actual theory of language acquisition and cognitive development, it is hard not to be skeptical of Davidson's views about the necessary relation between belief and language.

(4) *There are theoretically derived statistical laws of behavior.*

I have already argued that it is not just psychology but physics as well that at a fundamental level is based only on statistical laws. If there were more space, I would expand upon my argument to include the case of classical physics, once errors of observation are included in the theoretical analysis. But the real point is that fundamental physics in the latter half of the twentieth century, as opposed to the first half of the nineteenth century, is almost wholly statistical in character at a fundamental level. Sometimes, however, Davidson goes further, as, for example, in *PP* (p. 233) and in the subsequent discussion of his paper (pp. 239-44), to suggest that the kind of statistical laws that are characteristic of quantum mechanics cannot be achieved in psychology. As he puts it, 'The statistical laws of physics are serious because they give sharply fixed probabilities, which spring from the nature of the theory.' (A similar passage is found in *MM*, p. 250.) It is my claim that there are many examples of such serious statistical laws in psychology. Some of the best are to be found in mathematical theories of learning. This is not the place to present a detailed axiomatic formulation with derivation of theoretical statistical laws and accompanying evidence of their empirical correctness. However, I do want to make the point that the number of both theoretical and experimental papers on these matters is enormous, even though there is much that is still lacking to have the theory as adequate as we would like.

Although the subject-matter here is different from that of physics, the techniques of theoretical derivation of results and the use of general probabilistic tools of analysis are very similar. As an example, a simple model of all-or-none learning that may be thought of in terms of either conditioning or insight is easily described informally. Learning is a two-state Markov process depending on a single learning parameter c; if we call the states **U** for unlearned and **L** for learned, the transition matrix is

$$\begin{array}{c|cc} & \mathrm{L} & \mathrm{U} \\ \hline \mathrm{L} & 1 & 0 \\ \mathrm{U} & c & 1-c. \end{array}$$

The probability of a correct response in the state L is 1, and the probability of a correct response in the state U is p. It is also assumed that the initial state is U with probability 1. The mean learning equation giving the mean probability of a correct response p_n on trial n is then easily derived:

$$p_n = (1-c)p_{n-1} + c,$$

whence

$$p_n = 1 - (1-p)(1-c)^{n-1}.$$

All probabilities, for example, the distribution of last error, not just the mean learning curve, are a function of the two parameters of the model, c and p. Let **E** be the random variable for the trial of last error. Then the distribution of **E** is:

$$\mathrm{P(E\ \ = n)} = \begin{cases} bp & \text{for n} = 0 \\ b(1-p)(1-c)^{n-1} & \text{for n} > 0 \end{cases}$$

where

$$b = \frac{c}{1-p(1-c)}\ .^{5}$$

It is possible that Davidson will argue that this example falls outside of that part of psychology with which he is concerned, the part that makes essential use of intentional (and therefore intensional) concepts. The explicit classification is only hinted at in various passages by Davidson (for example, *PP* pp. 229-30, and discussion of *PP*, p. 240), and general reservations are not stated in *MM*, which was published before *PP* and *TT*. As I classify matters, the applications of the all-or-none learning model to concept learning of children fall within an intentional framework. In the first place, the concepts learned were elementary mathematical concepts that are a part of the curriculum the child is taught intentionally to learn and remember and that come to be a part of his beliefs about the world. Secondly, the experiments referred to were concept experiments in the following sense: no stimulus displays of sets, isosceles triangles, or the like were repeated, and thus no reductive theory of fixed stimulus-response connections could explain the learning. Thirdly, the theory does not postulate an observable point at which learning or insight occurs; only the pattern of responses is observable. The expected trial of learning, as opposed to trial of last (observable) error, is easily computed in theory

[5] Detailed application of this model, and more complicated extensions to the learning of elementary mathematical concepts by children, is given in Suppes & Ginsberg (1963), and Suppes (1965).

but it cannot be directly observed. Obviously this simple example does not postulate a very complex internal pattern in the learner, but it is easily extended to models that do (see, for example, Suppes, 1973).

(5) *Experimental tests of decision theory do not require an interpretation of speech.*

From the standpoint of quantitative theory in psychology, I find Davidson's remarks about decision theory puzzling. He mentions Ramsey's early work, casually describes an experiment of his own with Carlsmith (*PP*, pp. 235-6), and discusses briefly the transitivity of preference. The number of theoretical and experimental papers on these matters is very large. It is hard to think of a matter that has been more thoroughly investigated in various ways than the putative transitivity of indifference of preference. It is easy enough to agree with his remarks that we could improve decision theory by incorporating into it a theory of communication, but remarks of this kind about improvement can be made for almost any physical theory as well. The question is, rather, how he wants to evaluate scientifically the massive psychological literature on decision theory. It will be useful to focus on a single issue—Davidson's claim in *TT* (pp. 162-3)—that we cannot properly understand the choices an individual makes in expressing his preferences without relying on talk about these choices. Here is what Davidson has to say on the matter:

> What is certain is that all the standard ways of testing theories of decision or preference under uncertainty rely on the use of language. It is relatively simple to eliminate the necessity for verbal responses on the part of the subject: he can be taken to have expressed a preference by taking action, by moving directly to achieve his end, rather than by saying what he wants. But this cannot settle the question of what he has chosen. A man who takes an apple rather than a pear when offered both may be expressing a preference for what is on his left rather than his right, what is red rather than yellow, what is seen first, or judged more expensive. Repeated tests may make some readings of his actions more plausible than others, but the problem will remain how to tell what he judges to be a repetition of the same alternative. Tests that involve uncertain events—choices between gambles—are even harder to present without using words. The psychologist, skeptical of his ability to be certain how a subject is interpreting his instructions, must add a theory of verbal interpretation to the theory to be tested. If we think of all choices as revealing

> a preference that one sentence rather than another be true, the resulting total theory should provide an interpretation of sentences, and at the same time assign beliefs and desires, both of the latter conceived as relating the agent to sentences or utterances. This composite theory would explain all behavior, verbal and otherwise. (*TT*, pp. 162-3).

Davidson's claims in this passage raise important issues. To begin with, they seem to challenge the scientific methodology of a wide variety of psychological experiments. Concerning experiments involving human subjects, Davidson is certainly right in noting the extensive reliance on the use of verbal instructions. Does this mean that we must add a theory of verbal interpretation to each of the theories to be tested? In strictest terms, we could insist on such a theory, but exactly the same holistic problem arises in other sciences, such as physics. In the same spirit, we could insist on a theory of the actions of the physicist in performing an experiment. In this case, the actions the experimenter takes in preparing and using experimental apparatus correspond to the giving of verbal instructions to human subjects. It is part of the radical incompleteness of science that neither in physics nor in psychology do we ever satisfy the demands for the kind of composite theory including a full interpretation of instructions given to subjects or of actions taken with physical apparatus that Davidson seems to want. It is easy enough to agree with him that having such theories would be most desirable. But this will be a case of science made in heaven and not on earth.

Davidson also argues for the necessity of a theory able to interpret a subject's utterances about his preferences. He gives the example of a man who has taken an apple rather than a pear, but we cannot really tell, Davidson says, whether he is expressing a preference for what is on his left or what is on his right, what is red rather than yellow, or what. This problem is not in the least special to agency or psychological experiments. It is a standard problem of experimental design. If I have a hypothesis that a certain force is moving particles that are to be observed in a Wilson cloud chamber, I have exactly the same problem of eliminating other causes in order to give a univocal interpretation to the experimental results. I see nothing special about the case of preference. This is exactly what the subject of experimental design is about, and it is one of the marks of scientific progress in the twentieth century to recognize the need for and to have developed a theory of experimental design to disentangle the ambiguities of interpretation that Davidson poses, although I would not, of course, claim that we are always successful. We can bring the matter closer to psychology by examining the very extensive literature on preference in animals. If we took Davidson's arguments literally, we would not be able to make inferences of a definite kind about the preferences of animals (for example, for kinds of food, various solutions of sugar, etc.) because we are not able to relate the agent or subject to utterances, potential or actual. I certainly agree with Davidson about the importance of speech and its central role in understanding many kinds of decisions. What I cannot accept and do not believe is correct is his insistence on the necessity of tying the theory of decision and the theory of interpretation so closely together. It may be that he wants to make the more reasonable claim that for a certain important class of decisions a theory of interpretation of speech is necessary. In the passage cited and in other places he does not put such qualifications, and in his discussion of the question of whether animals can have beliefs, he clearly moves in the other direction. I am puzzled by how he would therefore want to interpret the vast literature on learning and preference in animals.

Finally, there is another, quite different point I want to make. Even in the case of complex and highly significant decisions, I am skeptical of an individual's ability to verbalize the basis for his choices. It seems to me that decisions we make about a variety of important matters are marked just by our inability to give anything like adequate explanations of why we have made the choices that we have made. To hold otherwise is a fantasy of rationality. If I am at all near the mark on this point, it is another reason for separating the theory of decision and the theory of how we talk.

I have been critical of various arguments of Davidson's that seem to raise important issues and yet have not been given by him in sufficient detail to be considered conclusive. Indeed, in some cases it seems to me his arguments move in a direction that is philosophically or scientifically mistaken. On the other hand, I want to stress my agreement with much of what Davidson says. His focus on the need for a general theory of desires, beliefs, and actions, and for a general theory of how we talk, rightly emphasizes matters that should be central to psychology but do not yet have a proper scientific foundation.

28

CURRENT DIRECTIONS IN MATHEMATICAL LEARNING THEORY

I have organized this article into two parts. In the first part I survey a number of different current trends in mathematical learning theory, with some attempt also to give some background of the developments leading up to them. In this part the main topics that I cover are stimulus-response theory, language learning, formal learning theory, and a group of related approaches I have entitled perceptrons, cellular automata, and neural networks. (The survey given here extends considerably the one given in Suppes, 1977a.)

In the second part I return to some of my own earlier work on stimulus-response theory of finite automata and give an extension to universal computation via register machines, rather than Turing machines. In this context I also discuss the feasibility of applying these theoretical ideas directly to actual learning situations.

1. GENERAL SURVEY

Stimulus-response theory. For the period running from the late 1930s to the late 1960s, that is, roughly a period of 30 years, the dominant theo-

*Reprinted from, *Mathematical psychology in progress* (ed. by E. E. Roskam), 1989, pp. 3–28. Berlin, Heidelberg: Springer Verlag.

retical view of learning was some variant of stimulus-response theory. We could, of course, begin earlier with the work of Thorndike, but for the period to which I am referring we can start with the papers of Clark Hull, and especially his *Principles of Behavior* (1943). On the other hand, Hull's theory does not have a genuine mathematical feel about it. It is impossible to make nontrivial derivations leading to new quantitative predictions of behavior. This is so in spite of the valiant attempts to formalize Hullian theory (Hull et al., 1940). In my judgment the first serious paper that had an impact in the history of mathematical learning theory was William K. Estes' article "Toward a Statistical Theory of Learning" (1950). Estes presented in statistical learning theory a theory that has the kind of formulation that we expect of theories in physics. Nontrivial quantitative predictions can be made, and especially we can vary experimental conditions and derive new predictions. Another early and important publication was the 1955 book of Robert Bush and Frederick Mosteller, *Stochastic Models for Learning.* In the period roughly from 1955 to 1970, a large number of additional theoretical and empirical studies appeared and I will not attempt to survey them here. I do want to mention my own 1969 article, "Stimulus-Response Theory of Finite Automata," because in the second part I will return to the framework of this article and extend it to register machines, thereby establishing connection with some of the other directions characteristic of mathematical learning theory in the last several decades. The main theorem of the 1969 article is the following:

THEOREM 1. *Given any finite automaton, there is a stimulus-response model that under appropriate learning conditions asymptotically becomes isomorphic to the finite automaton.*

There are one or two features of stimulus-response theory as developed in the period I have described that have not been really satisfactorily replicated in the two decades since. It is important to recognize that progress has not been uniform in this respect. The most important of these features is the one mentioned above, the ability to make new predictions based upon change in experimental parameters. What is important about these predictions, moreover, is that the predictions of stochastic learning models derived from stimulus-response theory were detailed in character compared to the often qualitative or single predictions made from learning theories derived from the current fashion in cognitive psychology. To give some sense of this, I cannot resist showing some results on stimulus-response theory for a continuum of responses. Figure 1 shows a comparison between an observed response histogram conditioned upon a preceding reinforcement with a corresponding predicted density for a

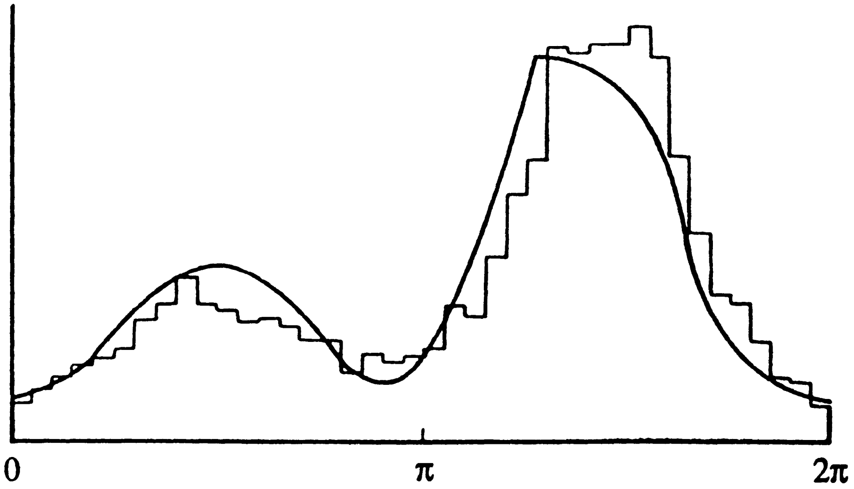

Figure 1. Observed response histogram conditional upon preceding reinforcement with corresponding predicted density.

continuum of responses with noncontingent bimodal reinforcement distribution. Details can be found in Suppes, Rouanet, Levine, and Frankmann (1964). (For a recent application of these ideas to learning by robots, see Crangle and Suppes, 1989b.) The derivation of this highly nonsymmetric curve is very much in the spirit of the kinds of derivations which one makes in physics and which work so well in stochastic learning models. The weakness of such models is that they work well in narrowly defined experimental settings, which I must say, by the way, is true of very many physical theories as well. The moral of the story is that in learning theory as in other subjects one cannot have one's cake and eat it too. If one wants precise mathematically derived predictions, then the experimental situations will probably be relatively narrowly circumscribed. On the other hand, if one wants to deal with a wide range of significant phenomena the predictions will not be nearly as satisfactory.

Mathematical models of language learning. Although the theory of language learning in general form has a long history, we can date the mathematical theory in the modern period from an important paper of Gold (1967). He established the following important theorem.

THEOREM 2 (Gold). *Regular or context-free classes of grammars are not text-learnable.*

By *text-learnable* is meant that just by being presented instances of text the grammar as such can be learned, that is, asymptotically identified. Note that the sense of learnable in this theorem is a very weak sense. One has in mind that the grammar cannot be learned even with an infinite number of trials. On the other hand, on the assumption that one could ask an informant whether something were grammatical in the language being spoken and therefore according to the grammar which was to be learned, Gold proved the following theorem.

THEOREM 3 (Gold). *Regular or context-free classes of grammars are informant-learnable.*

It is important to note, of course, that Gold's framework is completely nonpsychological and is based really just on formal properties of regular and context-free grammars.

The most notable effort to develop a mathematical but psychological theory of language learning is to be found in the various publications of Kenneth Wexler and his associates. Perhaps the first article to be mentioned is Hamburger and Wexler (1973), in which they study the identifiability of a class of transformational grammars, and their 1975 article on a mathematical theory of learning transformational grammar. Here I shall refer especially to the large book of Wexler and Culicover (1980), entitled *Formal Principles of Language Acquisition.* The general idea of their theory is that when one is given surface data, for example, spoken sentences, then each surface sentence is paired with a base phrase marker and this structure is then passed to a learning mechanism. The intuitive idea is that every pair (b, s) in a certain range has a fixed probability greater than zero of appearing at time t, and that this probability of appearing at time t is bounded away from zero independent of the past history of the system. Exactly which pairs do show up, that is, what the theory permits as possible transformations, is a matter for linguistic analysis, to which we turn in a moment. The learning mechanism is a hypothesis formation procedure that is familiar from many other contexts. At any particular time t, the state of the learner is represented by a finite set of transformations, and on each "trial" the learner is presented with a pair (b, s). In response to the pairing, that is, the given phrase structure of grammars is used to decide if it is correct and thereby make no change, or if it is incorrect try another hypothesis. Such learning only from errors is a familiar mathematical model of learning. So the learning mechanism itself is simple, easy to describe, and very much in the tradition of earlier mathematical models of learning.

Notice that knowledge of the base structure is assumed in Wexler and Culicover's theory of learnability. Moreover, this base structure is the carrier of meaning. This assumption is certainly psychologically unrealistic but can be accepted in the context of their theory in order to permit concentration on the problem of learning a transformational grammar. But what is actually going on in a young child is undoubtedly very much more complicated than such a simple theory postulates. What is complicated, and I shall not attempt to describe in detail here, is the set of transformations derived from the theory of transformational grammars. Wexler and Culicover impose five important structural restrictions on transformations, which can be justified linguistically. The five basic restrictions that they impose on the class of transformational grammars can be described in the intuitive terms I used in my original review of the book (Suppes, 1983).

1. The *freezing principle* asserts that if a transformation changes the structure of a node so that that part of the base structure is no longer a base structure (i.e., able to be generated by the context-free grammar of the base), then no transformations may be applied to subparts of the structure of the node. (The intuitive picture here is of a grammatical tree structure, and the "structure" of a node refers to that part of the tree lying below the node.) For example, if we applied a passive transformation (I am not claiming such transformations are psychologically or linguistically sound) to *John who loved Mary loved Jane* to obtain *Jane was loved by John who loved Mary,* we could not then apply a transformation to the subordinate relative clause.

2. The *binary principle* restricts transformations to applying to constituents that cut across more than two embedded sentences in the base structure. Primarily because of this principle, Wexler and Culicover are able to prove that the learner need never encounter base sentences more complex than having two stages of embedding. Thus, base sentences of the complexity of the familiar nursery rhyme *This is the dog that worried the cat that killed the rat that ate the malt that lay in the house that Jack built* need not be encountered.

3. The *raising principle* asserts that if a node is raised, a transformation cannot be applied to a node beneath this node. For example,

consider the sentence *John believes that the man who shot Bill loves Mary.* By raising we obtain *John believes the man who shot Bill to love Mary,* the noun phrase *the man who shot Bill* has been raised from subject of the complement clause to object of the main verb, and by the raising principle no transformation can be applied to the relative clause of this noun phrase.

4. The *principle of no bottom context* is rather technical, and I shall not try to illustrate it here. What it does is rule out lower structures that overly determine whether a transformation at a higher level fits exactly and is thus applicable.

5. The *principle of the transparency of untransformable base structures* is also technical. It asserts that base structures that cannot be transformed must turn up in the surface structure and thus be transparent.

THEOREM 4 (Wexler and Culicover). *With restriction of input to sentences satisfying the binary principle stated above, a transformational grammar also satisfying the other four principles listed above may be asymptotically learned with probability one.*

I have already remarked in several places on criticisms of the theory of the kind that are standard of mathematical theories of learning, namely, simplifications that all recognize are simplifications. There is another point of the Wexler and Culicover work that needs remarking of a different sort. They do not offer any evidence about the rate of learning. It could very well be that, on the assumption that sentences appear once a minute, it would still take several hundred years to learn a working grammar. In fact, the amount of time required could really be astronomical. It is an important aspect of theories of this kind to get, even if necessary by simulation, some idea of the rate of learning. Without doubt, the detailed features of the transformational grammar will have by far the largest impact on the rate of learning. Of course, by putting in hierarchical principles restricting in a given situation the transformations available, one can see how it would speed up very much the rate of learning and it may be by efforts of this kind a reasonable rate of learning could be achieved. What is probably needed in the tradition of this work at the present time is to get closer to some actual experiments—or actual data perhaps even better—of children's language learning, but the difficulties of testing quantitative theories in this fashion are also well known.

To some extent, this moving closer to the data is characteristic of the theory set forth in Pinker (1984). Also of interest is the fact that Pinker builds his theory around lexical functional grammars (Kaplan and Bresnan, 1982). Lexical functional grammars represent one of the new generalized phrase-structure grammars, which currently seem very promising from a linguistic standpoint and therefore represent a better target choice than the transformational grammars used by Wexler and Culicover. On the other hand, as Pinker explicitly states, he has not attempted to give a formal theory of language learning, so that consequently his main efforts fall outside the framework of this paper. His book is full of interesting remarks about problems of theorizing in this area and also about a great variety of psycholinguistic experimental and naturalistic data. But it would at the present time be extremely difficult to formalize his ideas as formulated in the book, and the standard questions we would want to ask of such a theory, as, for example, computations about rates of learning, etc., to show practical feasibility, are simply out of the question.

Formal learning theory. Based on the kind of idea characteristic of Gold's important early work (1967), a discipline known as formal learning theory has developed. The mathematical tools are essentially ideas of computability and recursion. There is no use of probabilistic notions and no appeal to actual empirical data. Thus, for example, the processes being modeled are not really actual learning processes, as exhibited in animals, persons, or even practical computer programs. The theorems mainly consist of possibility and impossibility theorems of the kind already exemplified in Theorems 2 and 3, due to Gold. The most recent summary of work in this field is the 1986 book *Systems That Learn* of Osherson, Stob, and Weinstein. Various combinations of these three authors have published a number papers as well. I will not try to summarize the variety of results that are proved but will state only one typical theorem that I think gives a sense of the kinds of results established. It is important to remember that this theorem is typical—it is not some conclusion of a large buildup.

THEOREM 5 (Osherson, Stob and Weinstein). *A learning strategy with limited memory restricts the set of languages that can be learned.*

Sketch of Proof. Let the memory for sentences of the strategy φ be only for the sentence currently being processed. Let L be the set of languages consisting of

$$\begin{aligned} L &= \{(0,i)|i \in N\}, \\ L_j &= \{(0,i)|i \in N\} \cup \{(1,j)\}, \\ L'_j &= \{(0,i)|i \neq j, i \in N\} \cup \{(1,j)\}. \end{aligned}$$

When φ first sees $(1,j)$ for some j, φ cannot remember whether it previously saw $(0,j)$ or not, and so cannot distinguish between L_j and

L'_j. This argument may easily be extended to a φ having any fixed finite memory.

Osherson, Stob, and Weinstein have a lot of informal remarks about human learners and language learning as well as about children's learning of first language, but it is obvious enough that their approach could not possibly supply anything like an empirically adequate theory of language learning. They establish within the framework of recursion theory a variety of positive and negative results but they do not move from general theorems of a recursive sort even to theorems about feasibility. Typically it has been important in the recent literature on complexity in computer science, for example, to distinguish a recursive result from special cases that are also feasible. The typical case is that a decision procedure that is exponential in terms of some essential parameter is recursive but not feasible. In order to be feasible it must be no more than polynomial in number of steps in the parameter. A typical example of this would be Tarski's well-known decision procedure for elementary algebra. It is known to be strongly exponential in the length of the algebraic formula whose validity is being decided. Feasible procedures can be given only for special subsets.

In spite of the general references to children, for example, and other aspects of human learning, there is no direct analysis of data of any kind and it would not seem appropriate with the self-imposed restrictions of the authors. Along with their absence of any detailed complexity results there are also no concrete or numerical results on the rate of learning of the various learning strategies they investigate. The results are either negative, for example, something cannot be learned, or the results are asymptotic. The concrete and positive study of learning must lie elsewhere, even though the kinds of results given can be useful in establishing the boundaries of substantive learning theories.

Also to be mentioned is the sophisticated independent work of Ehrenfeucht and Mycielski (1973a, 1973b, and 1977). The general framework is similar to that of Osherson, Stob, and Weinstein, but Ehrenfeucht and Mycielski introduce different conceptual ideas in order to characterize asymptotically learnable functions. Equally important in this connection is the independent work of Pentti Kanerva (1988), who goes further than Ehrenfeucht and Mycielski on the difficult problems of the organization of memory for formal learning theories or for related computing performance issues.

Perceptrons. I have grouped under this heading a great variety of work that I cannot attempt to survey in any depth but all of which needs at least some mention. The work on perceptrons goes through a line beginning with McCulloch and Pitts (1943), Rosenblatt (1959), and Minsky and Papert (1969). Minsky and Papert were quite successful in setting perceptron theory in a mathematical framework in which useful results could be proved. Their most important results were negative, showing that simple perceptrons of the kind mainly discussed in the prior literature were not capable of learning some quite simple geometrical concepts. I will not enter into the details of these negative results but define the concepts necessary to state the main positive result, the perceptron convergence theorem, particularly in the form given by Minsky and Papert.

The conceptual apparatus is of the following sort. There is a set D of perceptual displays and a subset G that we want the device to learn to select correctly, that is, to identify correctly membership or nonmembership in G. More particularly, on each trial the device responds *yes* to the presented display d if it classifies d as a member of G; otherwise it responds *no*. If the answer is correct, the device receives positive reinforcement (e_1), and, if the answer is incorrect, it receives negative reinforcement (e_2). Notice that the reinforcement is symmetric with respect to yes-no responses and depends only on the correctness or incorrectness of the response. This type of reinforcement is what I shall term standard nondeterminate reinforcement. Later on I shall discuss more complex and more informative reinforcement structures. The heart of the idea of the perceptron is that there is a set ϕ of elementary predicates. Each of these predicates is assigned a weight, and the perceptron combines in linear fashion the weighted 'answers' of the predicates to answer more complex questions. We can state the basic definition and the basic theorem without specifying more exactly the character of the elementary predicates because the results are relative to the set ϕ of predicates. For notational purposes, let A be a k-dimensional vector of real numbers that gives the weights assigned to each of the k-elementary predicates with a_i being the weight assigned to elementary predicate ϕ_i. The set ϕ can also most easily be represented as a k-dimensional vector of the k-elementary predicates. For each perceptual display d in D, $\phi_i(d)$ has the value 1 if d has the property expressed by ϕ_i and 0 otherwise. We can thus use standard inner product notation for vectors so that in place of $\sum a_i\phi_i(d)$ we can write $A \cdot \phi(d)$. It is understood that the response of the perceptron learning model is *yes* if this inner product is greater than 0 and *no* otherwise. To refer to a particular trial, the vector A_n of coefficients can also be referred to as the state of conditioning of the perceptron learning model at the beginning of trial n, and d_n is the object presented on trial n.

In the present context, finally the sample space X consists of all possible experiments, with each experiment, of course, being a sequence of trials. A trial in the present case is simply a triple (A, d, e), where A is the state of conditioning as described already, d is a perceptual display that is a member of D, and e is a reinforcement.

We thus have the following definition.

DEFINITION 1. *A structure* $\mathcal{A} = (\mathcal{D}, \phi, \mathcal{E}, \mathcal{X})$ *is a* perceptron learning model *if and only if the following axioms are satisfied for every sequence of trials in X.*

(i) *If* e_1 *occurs on trial* n, *then* $A_{n+1} = A_n$.

(ii) *If* e_2 *occurs on trial* n *and* $A_n \cdot \phi(d_n) \leq 0$, *then* $A_{n+1} = A_n + \phi(d_n)$.

(iii) *If* e_2 *occurs on trial* n *and* $A_n \cdot \phi(d_n) > 0$, *then* $A_{n+1} = A_n - \phi(d_n)$.

Note that the main feature of the learning of perceptrons is that learning occurs only when an error is made, that is, when an e_2 reinforcement occurs. The vector expressing the state of conditioning changes in one of two directions, depending upon the type of error.

In terms of these concepts we can then state the standard theorem.

THEOREM 6 (Perceptron Convergence Theorem). *For any set* D *and any subset* G *of* D, *if there is a vector* A *such that* $A \cdot \phi(d) > 0$ *if and only if* $d \in G$, *then in any perceptron learning model there will only be a finite number of trials on which the conditioning vector changes.*

What is particularly important to note about this theorem is the ergodic property of convergence independent of the particular choice of weights A_1 at the beginning of the experiment. The finite number of changes also implies that the perceptron learning model will make only a finite number of mistakes. The hypothesis about G expressed in the theorem is equivalent to saying that G and its complement with respect to D are linearly separable.

The perceptron convergence theorem is asymptotic in character. The more important question of rate of learning is not settled by the theorem, although there exist some results in the literature that are of some interest. I say no more about perceptrons here because work on them in many ways converges into other work to be mentioned.

Cellular automata. To illustrate ideas, here is a simple example of a one-dimensional cellular automaton. In the initial state, only the discrete position represented by the coordinate 0 has the value 1. All other integer coordinates, positive or negative, have the value 0. The automaton can in successive 'moves' or steps produce only the values 0 or 1 at any given location. The rule of change is given by a function that depends on the value of the given position and the values for the simple case to be considered here of the adjacent values on either side. Thus, using a_i' as the value at site i, after applying the function for change, we can represent the rule of change as follows:

$$a'_i = \Phi(a_{i-1}, a_i, a_{i+1}). \tag{1}$$

Note that because of the restriction in the values at any site this function takes just eight arguments, the eight possibilities for strings of 0's and 1's. The automata being discussed here are at the very lowest rank on the number k of possible values: $k = 2$ is the number of possible values and $r = 1$ is the distance away. Updating may depend on adjacent values. The characterization here in terms of the infinite line can be replaced by a finite characterization, for example, in terms of a discrete set of points on a circle. The same kind of function as given in equation (1) applies.

Cellular automata have the following obvious characteristics:

Discrete in space

Discrete in time

Discrete state values

Homogeneous—in the sense all cells are identical

Synchronous updating

Deterministic rule of change corresponding to a deterministic differential equation of motion in the case of classical mechanics

Locality of change rule: the rule of change at a site depends only on a local neighborhood of the site

Temporal locality: the rule of change depends only on values for a fixed number of preceding steps—in the present example just one step.

The study of cellular automata by physicists has to a large extent been concerned with their usefulness as discrete idealizations of partial differential equations. Both computer simulation and mathematical analysis by use of cellular automata are simplifications of the difficult problems of the behavior of fluids and other complicated phenomena that in principle are governed by nonlinear differential equations.

The learning aspects of cellular automata have not been a real focus of study, but the closely related topic of self organization has been. The investigation of simple self-organizing phenomena goes back at least to the early work on chemical systems by Turing (1952). Even before that von

Neumann began lecturing and talking about cellular automata. It seems that the idea of cellular automata originated in conversations between von Neumann and Stanislaus Ulam (von Neumann, 1966) as idealized mathematical models of biological systems capable of self-reproduction. Von Neumann's construction of a self-reproducing automaton was one of the early conceptual successes. Von Neumann's two-dimensional cellular automaton construction consisted of a universal Turing machine embedded in a cellular array using 29 states per cell with 5-cell neighborhoods. Moreover, his automaton can construct in the cellular array any configuration of machine which can be described on its input tape. For this reason von Neumann's cellular automaton is called a universal constructor. Also important is the fact that transcription takes place: von Neumann's automaton also makes a copy of the input tape and attaches it to the constructed automaton.

THEOREM 7 (von Neumann). *There exists a self-reproducing finite cellular automaton.*

As would be expected, there are a large number of subsequent results about self-reproduction. A good review is given in Langton (1984), which includes a description of Codd's substantial simplification of von Neumann's construction (Codd, 1968), as well as his simplification of Codd. The earlier articles on self-reproduction are well represented in the volume edited by Burks (1970). The abstract theory of self-reproduction is especially well set forth in Myhill (1964), reprinted in Burks' volume.

Although there has been almost no literature on cellular automata directly focused on learning, there has been considerable interest in adaptive automata as models of biological systems. For general theory, see Holland (1975); for a review of what is known about emergent properties of random cellular automata, see Kauffman (1984) and references therein. Especially suggestive is the work of Burks and Farmer (1984) on the modeling of DNA sequences as automata. A recent encyclopedic review of automata and computer models in biology is Baianu (1986).

Because of their great definitional simplicity, cellular automata provide a good framework for contrasting determinism and predictability. Clearly, the nonprobabilistic cellular automata are deterministic in character, but predictability of their behavior is another matter. Knowing the transition rule for a cellular automaton and given an initial configuration, can one predict in closed form the configuration after n steps? In general, this is certainly not the case. Of course, we can get predictability in another way by directly simulating the automaton but this is not what we ordinarily mean by predictability. There is in fact a concept that is useful to introduce at this point. This is the concept of being *computationally* *irreducible* (Wolfram, 1985). A system is computationally irreducible if a prediction about the system cannot be made by essentially shorter methods than simulating the system or running the system itself. Wolfram (1986) has shown that the following $k = 2, r = 1$ cellular automaton generates highly complex sequences that pass many tests for randomness, in spite of its totally elementary character. The automaton is defined by equivalent equations, one in terms of exclusive *or* and the other in terms of mod 2 arithmetic.

$$a_i' = a_{i-1}XOR(a_iORa_{i+1})$$
$$a_i' = (a_{i-1} + a_i + a_{i+1} + a_ia_{i+1})\ mod\ 2$$

Another good example of a physical process that is computationally irreducible is the addition of a column of numbers on the fastest computer available. For numbers randomly chosen but that do not have some special tricky features, there is no shorter way to compute their sum than simply to run the algorithm of addition itself. Because of our great interest in predictability we often forget how many processes are not predictable in the sense that they can be predicted in advance of their actual running or direct simulation of running. Of course, we can learn many things about the behavior of a system, including the asymptotic behavior of a learning system, without having it be computationally reducible. A tantalizing conjecture is that the network of neurons in a human brain constitute in their complete computational capability a computationally irreducible system. The implications of this conjecture for having theories of learning or performance as detailed as one might ask for are quite pessimistic.

Neural networks. Just as cellular automata and perceptrons grow out of a common background, so do neural networks arise from the same background. Above all, neural networks are close to perceptrons in general conception. Neural networks differ from cellular automata in two essential respects. One is that updating is asynchronous, and, secondly, connections can be continually changed. (Cellular automata have been studied with these two changes in mind, but not extensively.)

Among the early work on neural networks we mention especially the voluminous research in various forms by Grossberg and his colleagues. Simply for reference I mention here Grossberg (1974, 1978, 1980, 1982). Grossberg's and his colleagues' research is concentrated on a psychophysiological theory of sensory processing and development. He gives an excellent overview of the theory in Grossberg (1978). As he puts it there, the theory is organized into four stages. The first stage is concerned with the question how fluctuating patterns of data are processed in cellular tissues so that problems of noise and dynamic range are solved. In particular, what are the mechanisms that keep the patterns of data from either being saturated or hopelessly confounded with noise? Stage two concentrates on the question how do persistently correlated presentations of two or more events yield an enduring reorganization of system dynamics. Grossberg's intention is to discuss problems at this stage, for example, classical conditioning, in a way that is fundamental in terms of schematic analysis of chemical and electrical properties of cellular structures. I emphasize the word schematic. It would be misleading to give a sense that he is dealing with the complete details of cellular dynamics.

The third stage concerns the question how sensory codes reorganize themselves or develop in response to environmental pressures. The theory at this stage particularly deals with the hierarchical organization of feature detectors, and in particular their interaction. Stage four deals with the fundamental question of how the information code can be held stable in an environment that is changing both internally and externally.

To give a sense of how Grossberg attacks these matters, a familiar and typical instance would be his analysis of adaptation and automatic gain control in on-center and off-surround networks which are hypothesized as schematic models of important aspects of the visual system. The models proposed are meant to account for psychologically known facts about perception of overlapping colored patches of light. Models are formulated in terms of population of cells which themselves have schematic properties in terms of activation decay. I hope it is clear that I am quite positive about the level of schematization that Grossberg is attempting. It is certainly out of range at the present time to try to use all the detailed information that is known about cell structure and, in particular, about the structure and function of neurons. How complicated the details are and how little we still know about function are well brought out in the chapter on the cerebral cortex by Crick and Asanuma in McClelland, Rumelhart, et al. (1986).

On the other hand, I will not try to review Grossberg's work in detail for another reason. He has concentrated on problems that are rather different from my focus here, namely, the computational power of learning models. There are certainly remarks in various publications that bear on this but he has mainly concentrated on specific network models of vision that account for visual illusions, on the development of feature detectors in the visual cortex, adaptive pattern classification, networks that account for psychological facts of olfaction, and especially the excellent recent monograph on the neural dynamics of ballistic eye movements (Grossberg & Kuperstein, 1986).

The surprising ability of neural networks to self-organize themselves to have emerging computational properties has recently been studied in simulations by Hopfield (1982) and Hopfield and Tank (1985). In the latter article it is shown how a neural network can be constructed to compute a surprisingly good solution of the traveling salesman problem, which is the hard combinatorial problem of finding the minimum-distance route a salesman should take in visiting n cities exactly once.

Within psychology the recent surge of interest in neural networks has been led by Rumelhardt, McClelland and the PDP Research Group (1986). Their massive and discursive two-volume treatise summarizes in one form or another most of the work done to date. What is especially interesting is the intellectual conflict that has already surfaced and is certain to continue between cognitive psychologists who are devoted to rule learning, and learning psychologists like Rumelhardt who do not believe that explicit rules, e.g., grammatical rules, play an important part in language learning or other cognitive domains. Chapter 18 of the work cited above—this chapter is by Rumelhardt and McClelland—provides a good example in its application of neural networks to learning the past tense of English verbs. Nothing like a complete sample of verbs is presented but enough is done to show what is a practical conceptual approach to developing a more extensive and detailed theory.

For an essentially negative critique of this work on verbs and related developments, see Pinker and Mehler (1988). The negative arguments by various cognitive psychologists in this volume are by no means conclusive, even though some excellent criticisms of neural networks, or connectionism as the theory is often called, are given. The important point from the standpoint of the present paper is that none of the criticisms by Pinker and others are precisely formulated from a mathematical standpoint. In other words, unlike other work cited here, no impossibility results are proved or even explicitly formulated.

There is a considerable development of formal theory in various other chapters of the work, with many of the ideas derived more from physics than psychology, even though some of the theorems state learnability results. An example is Chapter 6 by Smolensky on the foundations of harmony theory, a particular approach to information processing in dynamical systems, a topic close to some of the extensive work done by others on cellular automata. Smolensky contrasts his “subsymbolic” paradigm using concepts of activation, relaxation, and statistical correlation to the symbolic paradigm forcefully presented by Newell (1980), which emphasizes the central role of symbolic computation in cognitive science. This is the conflict mentioned above under another guise. The main characteristics of harmony theory, which is in its way a new descendant of older perceptron theories, are these: inference is through activation of schemata, stored knowledge atoms are dynamically assembled into context-sensitive schemata, schemata are coherent assemblies of knowledge atoms, harmony derives from making inferences that are consistent with the knowledge represented by the activated atoms, and the self-consistency of a possible state is assigned a quantitative value by a harmony function. What I have said here is only meant to be suggestive, and many of Smolensky's remarks are of the same sort. However, and I emphasize this point, in the appendix to the chapter he gives a straightforward mathematical development of harmony theory. It is just not feasible to state here in technical detail the three theorems on competence, realizability, and learnability.

2. LEARNING THEORY FOR UNIVERSAL COMPUTATION

Finite automata again. The main purpose of this section is to straighten out confusions that have been the sources of criticisms of Theorem 1 (Suppes, 1969b) on stimulus-response models of finite automata. I begin with a brief summary of this earlier work.

The central idea is quite simple—it is to show how by applying accepted principles of conditioning, an organism may theoretically be taught by an appropriate reinforcement schedule to respond as a finite automaton. When an automaton is presented with one of a finite number of letters from an input alphabet, as a function of this letter of the alphabet and its current internal state, it moves to another one of its internal states. In order to show that an organism obeying general laws of stimulus conditioning and sampling can be conditioned to become an automaton, it is necessary first of all to interpret, within the usual run of psychological concepts, the notion of a letter of an alphabet and the notion of an internal state. In my own thinking about these matters, I was first misled by the perhaps natural attempt to identify the internal state of the automaton with the state of conditioning of the organism. This idea, however, turned out to be clearly wrong. In the first place, the various possible states of conditioning of the organism correspond to various possible automata that the organism can be conditioned to become. Roughly speaking, to each state of conditioning there corresponds a different automaton. Probably the next most natural idea is to look at a given conditioning state and use the conditioning of individual stimuli to represent the internal states of the automaton. In very restricted cases this correspondence works, but in general it does not. The correspondence that turns out to work is the following: the internal states of the automaton are identified with certain responses of the organism.

I now turn to the discussion of Theorem 1. The most technically detailed and in many ways the most interesting criticism of this representation theorem has been by Kieras (1976). The intuitive source of Kieras' confusion in his claim that the theorem as stated is too strong is easy to identify. Because I identified the *internal states* of a given automaton with the *responses* of the representing stimulus-response model, Kieras inferred I had unwittingly restricted my analysis to automata that have a one-one correspondence between *their* internal states and responses. On the basis of this confusion on his part he asserts that the representation theorem is not correct as it stands.

My purpose now is to lay out this dispute in an explicit and formal way in order to show unequivocally that Kieras is mistaken and the representation theorem is correct as originally stated.

From a mathematical standpoint Kieras' mistake rests on a misunderstanding of representation theorems. The isomorphism of a representation theorem is a formal one. In the case of Theorem 1 above, the isomorphism is between the internal states of the automaton and the responses of the representing stimulus-response model. The Rabin-Scott definition of automata used in the 1969 article does not have an explicit response mechanism, but that this is a trivial addition to their definition is shown by the *general* definition of a sequential machine with output given by Harrison (1965, p. 294), who is referenced by Kieras and who acknowledges he is mainly following the terminology of Rabin and Scott (see p. 292). An automaton or sequential machine with output is for Harrison just an automaton in the sense of Rabin and Scott with the additional condition that the set F of final states are those "giving a one output." As Harrison and others have remarked, a restriction of output to 1's and 0's is no restriction on the generality of the sequential machine.

The addition of this output apparatus to the formal definitions I gave in the original article is trivial. We just pick two responses r_0 and r_1 not used to represent internal states, but one of them, say r_0, represents 0 and the other 1. Whenever the machine is in an internal state that is not a final state but a response is required, it outputs r_0. When it is in a final state it outputs r_1 . To modify Definition 1 to take account of these output responses is easy. I note once again that the two output responses are in no way intended to correspond to internal states of the automata being represented. Other responses of the stimulus-response model represent the internal states. I emphasize, also, that this modification of adding output responses would not be correcting an error in Theorem 1 but would only be providing an additional closely related result.

Register machines. Another misconception in the literature is that stimulus-response theory can only deal with machines that have the power of finite automata. The purpose of this section is to show that this is not the case by giving the construction of register machines, which are equivalent to Turing machines, corresponding to that for finite automata. The development here extends and modifies substantially that in Suppes (1977b). To give the results formal definiteness, we shall develop a learning theory for any partial recursive function. Such functions can be defined explicitly in a fairly direct way but we shall not do so here. I shall rely upon the fact that partial recursive functions are computable functions. We then use the basic theorem in the literature, whose technical framework we shall expand upon somewhat later, that any function is partially recursive if and only if it is computable by a register machine or, equivalently, by a Turing machine. The concept of a register machine used here was introduced by Shepherdson and Sturgis (1963). The reason for using register machines rather than Turing machines is that their formal structure is simpler. For example, the proof of equivalence between a function being a partial recursive function and being computable by a register machine is much simpler than the corresponding proof for Turing machines. First, let me recall how simple a classical register machine for a finite vocabulary is. All we have is a potentially infinite list or sequence of registers, but any given program uses only a finite number. Exactly three simple kinds of instructions are required for each register. The first is to place any element of the finite vocabulary at the top of the content of register n; the second is to delete the bottommost letter of the content of register n if the register is nonempty; because any computation takes place in a finite number of steps, the content of any register must always be finite in length. The third instruction is a jump instruction to another line of the program, if the content of register n is such that the bottommost or beginning letter is a_i; in other words, this is a conditional jump instruction. Thus, if we think of the contents of registers as being strings reading from left to right we can also describe the instructions as placing new symbols on the right, deleting old symbols on the left, and using a conditional jump instruction in the program when required.

It is straightforward to give a formal definition of programs for such an unlimited register machine, but I delay this for the moment. It is clear that a program is simply made up of lines of instructions of the sort just described. The potentially infinite memory of an unlimited register machine both in terms of the number of registers and the size of each register is a natural mathematical idealization. It is also possible to define a single-register machine with instructions of the kind just stated and to show that a single register is also adequate.

An important point about the revision of stimulus-response theory given here is that the internal language used for encoding stimulus displays is all that is dealt with. In other words, in the present formulation of the register-machine theory I shall not enter into the relation between the set of external stimuli and the encoding language, but deal only with the already encoded representation of the display. This level of abstraction seems appropriate for the present discussion but of course is not appropriate for a fully worked out theory. It is a proper division of labor, however, with the proper modularity. I am assuming that the sensory system passes to the central nervous system such encoded information, with the first level of encoding taking place well outside the central nervous system. Thus, in one sense the concept of stimulus becomes nonfunctional as such, but only because the encoding is already assumed. It is obvious enough that no serious assumptions about the actual perceptual character of stimuli is a part of classical S-R theory. Secondly, the concept of a program internally constructed replaces the direct language of responses being conditioned to stimuli. A natural question would be why not try to give a more neural network or hardware version of this construction. Given how little we know about the actual way in which information is transduced to the central nervous system and then used for encoding and programming, it seems premature, and in fact may well be premature for much longer than many of us hope, to try to move to any hardware details. Certainly what does seem to be the case is that there is internal programming. I am not suggesting that the abstract simple theory of a register machine catches the details of that internal programming—it is only a way of representing it—, and it is a matter for detailed additional theory to modify the abstract representation to make it more realistic.

On the other hand, without giving anything like a detailed neural analysis, the register-machine programs can be replaced by computationally equivalent stimulus-response connections, but without further specification such postulated S-R conditioning connections are no more concrete, i.e., closer to empirical realization, than the register-machine programs. It seems to me that it is therefore better to think of the programs as being realized by neural "hardware" we cannot presently specify. What is presented in the remainder of this section is formally adequate, but can surely be improved upon in many ways either to more closely imitate the learning of different organisms or to make machine learning more efficient. Moreover, given some feature coding of presented stimuli, there is every reason to think that to any software program there is a corresponding neural net, and vice versa, for solving a particular class of problems with essentially the same rate of learning. But this likely equivalence cannot be pursued further here.

To make matters more explicit and formal but without attempting a complete formalization, I introduce the following definitions. First, $\langle n\rangle$ is the content of register n before carrying out an instruction; $\langle n'\rangle$ is the content of register n after carrying out an instruction. Second, a *register machine* has (1) a denumerable sequence of registers numbered $1, 2, 3, \ldots$, each of which can store any finite sequence of symbols from the basic alphabet V, and (2) three basic kinds of instructions:

(*a*) $P_N^{(i)}(n)$: Place a_i on the right-hand end of $\langle n\rangle$.
(*b*) $D_N(n)$: Delete the leftmost letter of$\langle n\rangle$ if $\langle n\rangle \neq 0$.
(*c*) $J_N^{(i)}(n)[q]$: Jump to line q if $\langle n\rangle$ begins with a_i

If the jump is to a nonexistent line, then the machine stops. The parameter N shown as a subscript in the instructions refers to the set of feature registers holding sensory data and not used as working computation registers. (This point is made more explicitly in the definition given below.)

A *line* of a program of a register machine is either an ordered couple consisting of a natural number $m \geq 1$ (the line number) and one of the instructions (*a*) or (*b*), or an ordered triple consisting of a natural number $m \geq 1$, one of the instructions (*c*), and a natural number $q \geq 1$. The intuitive interpretation of this definition is obvious and will not be given.

A *program* (of a register machine) is a finite sequence of k lines such that (1) the first number of the i^{th} line is i, and (2) the numbers q that are third members of lines are such that $1 \leq q \leq k+1$. The parameter k is, of course, the number of lines of the program. I shall also refer to programs as *routines.* How a register machine *follows* a program or routine is intuitively obvious and will not be formally defined. *Subroutines* are defined like programs except (1) subroutines may have several exits, and (2) third members of triples may range over $q_1, \ldots, q_k$, these variables being assigned values in a given program.

I shall not give the formal definition of a partial recursive function defined over the alphabet V. It is any intuitively computable function. Given V, the finite vocabulary, then, as usual in such matters, $V*$ is the set of finite sequences of elements of V; in the present context, I shall call the elements of $V*$ *feature codings.* Let f be a function of n arguments from $V * x \cdots xV*$ (n times) to $V*$. The basic definition is that f is computable by a register machine if and only if for every register x_i, y and N with $y \neq x_i$ for $i = 1, \ldots, n$ and $x_i, \ldots, x_n, y \leq N$ there exists a routine $R_N(y = f(x_1, \ldots, x_n))$ such that if $\langle x_1\rangle, \ldots, \langle x_n\rangle$ are the initial contents of registers $x_1, \ldots, x_n$ then

(1) if $f(\langle x_1\rangle, \ldots, \langle x_n\rangle)$ is undefined the machine will not stop,
(2) if $f(\langle x_1\rangle, \ldots, \langle x_n\rangle)$ is defined, the machine will stop with $\langle y\rangle$, the final content of register y, equal to $f(\langle x_1\rangle, \ldots, \langle x_n\rangle)$, and with the final contents of all registers $1, 2, \ldots, N$, except y, the same as initially.

I turn now to the axioms for register learning models that in a very general way parallel those given for stimulus-response models with nondeterminate reinforcement in Suppes and Rottmayer (1974). I axiomatize only the model, and not the full probability space that serves as a formal framework for the learning trials. Extension to the latter, possibly via random variables and leaving the probability space implicit, is straightforward but tedious.

The axioms are based on the following structural concepts:

(i) the set R of registers,
(ii) the vocabulary V of the model,
(iii) the subset F of feature registers,
(iv) the subset C of computation registers,
(v) the subset Rp of response registers,
(vi) the working memory WM,
vii) the long-term memory LTM,
(viii) the responses r_0 and r_1 ,
(ix) the real parameters ρ and c.

It will perhaps be useful to say something briefly and informally about each of the primitive concepts. The feature registers in F just encode the features of the presented stimulus. This encoding and computation as well is done by using the vocabulary V. The computer registers in C are working registers available as needed for computation. The working memory WM stores programs being constructed. For simplicity here I shall assume there is only one such memory, but clearly this is too restrictive for general purposes. The long-term memory LTM is where programs that are found by repeated trials to be correct are stored.

One distinction is essential between the two memories and the registers. The memories store the program, so to the feature vocabulary $v_1, \ldots, v_n$ in V is added notation for the three types of instruction: P for placing or adding on the right, D for deleting on the left, and J for a jump instruction. V must also include notation for referring to registers used and to program lines. For the purpose I add the digit 1 (thus $2 = 11$, $3 = 111$, etc.), the most rudimentary counting notation.

The set Rp of response registers is also here for simplicity assumed to be a singleton set. This register corresponds in the general register machine characterized earlier to be the register that holds the value of the partial recursive function being computed. Here also I make an inessential simplifying assumption, namely, that learning will be restricted to concept learning, which is in principle no restriction on the set of computable functions. In the present case, given that the program is completed, if the register is cleared, the response is r_0, which means that the stimulus displayed—whose features are encoded in F—is an instance of the concept being learned, and if the register is not empty the response is r_1 , which means the stimulus presented is not an instance of the concept. Moreover, if the program at any step is halted before completion, the response is r_0 with guessing probability ρ, and r_1 with probability $1 - \rho$.

The two real parameters ρ and c enter in the axioms in quite different ways. As just indicated, ρ is the response guessing probability, and c is the constant probability of stopping construction of a program. These parameters, and others introduced implicitly in the axioms, are surely context dependent, and will naturally vary from task to task.

As formulated here, each line of a program is run as it is selected for the program construction and placed in working memory (WM). A program is transferred to long-term memory (LTM) only when it is completed and is successful in correctly identifying an instance of the concept being learned. The mechanism of completely erasing a constructed program that is in error is too severe, but is a simplifying assumption that holds for some animal learning, e.g., the all-or-none elimination of habituation in aplysia by sensitiving stimuli (Kandel, 1985).

The three types of register-machine instructions—adding on the right, deleting on the left, or conditional jump—mentioned earlier are modified in one respect. To jump to a nonexistent line and thereby halt the program, rather than jumping to $m + 1$ where m is the number of lines, the jump is to 0, which is a possible number for no line. The reason for this change should be apparent. As the program is probabilistically constructed line by line by the learning model, there is no way of knowing in advance how long the program will be. So it is convenient to have in advance a fixed "place" to jump to in order to halt the program.

DEFINITION 2. *A structure* $\Re = (R, V, F, C, Rp, WM, LTM, r_0, r_1, \rho, c)$ *is a* register learning model for concept formation *if and only if the following axioms are satisfied:*

Register Structure Axioms

R1. *The subsets F, C, and Rp of registers are nonempty and pairwise disjoint.*

R2. *Subsets F and Rp, and the set V are finite and nonempty.*

R3. *Each register in R can hold any word of V_1^*, i.e., any finite string of elements of $V_1 = V - \{1, P, D, J\}$.*

Stimulus Encoding Axiom

D1. *At the start of each trial, the stimulus presented is encoded as having features $\langle f \rangle$ in the registers f* of *F.*

Program Construction Axioms

P1. *If at the start of the trial, the LTM is nonempty, no program construction occurs.*

P2. *Given that LTM is empty:*

(i) *With probability $c, 0 < c < 1$, construction of the program in WM terminates after each line, independent of the trial number and any preceding subsequence of events;*

(ii) *Given that a line is to be added to the program, the probability of sampling an instruction of any type with any argument is positive, independent of the trial number and any preceding subsequence of events; in the case of the line number n to which a jump is to be made the probability is geometrically distributed.*

Program Execution Axioms

E1. *If LTM is nonempty, the contents are copied into WM, and then the program is executed.*

E2. *If LTM is empty, then a program is constructed probabilistically, line by line according to Construction Axioms P1 and P2, and is executed as each line is constructed.*

E3. *When a jump instruction is executed, there is a fixed positive probability the program is halted after one step, with this probability being independent of the trial number and any preceding subsequence of events.*

Response Axioms

Rp1. *If when the program is complete, register Rp is empty, the response is r_0.*

Rp2. *If when the program is complete, register Rp is nonempty, the response is r_1.*

Rp3. *If the program is halted by Axiom E3, response r_0 is made with guessing probability ρ, and response r_1 with probability $1 - \rho$, the probability ρ is independent of the trial number and any preceding subsequence of events.*

Program Erasure Axioms

Er1. *If positive reinforcement occurs at the end of a trial, the program in WM is copied in LTM if LTM is empty.*

Er2. *If negative reinforcement occurs at the end of a trial, the program in WM is erased and so is the program in LTM if it is nonempty.*

A few of the axioms require comments that were not made earlier in the informal discussion. The probabilistic program construction axiom P2 is similar to a stimulus sampling axiom which guarantees accessibility for conditioning of all relevant stimuli. Axiom P2 is obviously formulated in such a way as to bound sampling probabilities away from asymptotically approaching zero except in the case of the geometric distribution for sampling line numbers. The stopping probability required in program execution axiom E3 is required in order to prevent staying with programs that generate infinite loops. Finally, the informal concept of reinforcement used in the program erasure axioms has an obvious meaning and is easily formalized. Positive reinforcement here just means that the concept classification of a stimulus by the response r_0 or r_1 is correct, and negative reinforcement that it is incorrect. Obviously, more informative reinforcement methods can and are widely used in learning and without question facilitate the speed of learning. More is said on this point in the final remarks on hierarchical learning.

On the basis of the axioms stated above we may prove an asymptotic learning theorem corresponding in a general way to Theorem 1 for stimulus-response models.

THEOREM 8. *Let f be any partial function of n arguments over the finite alphabet V and having just two values in V. Then f is a partial recursive function if and only if f is asymptotically learnable with probability one by a register learning model $\mathfrak{R}$ of concept formation.*

Proof. Let $\wp$ be a program for $\mathfrak{R}$ that computes f. We know there must be such a program by virtue of the fact that a function f over a finite alphabet is partial recursive if and only if it is computable by a register machine. Furthermore, given a definition of f we have a constructive method for producing $\wp$. Our objective is to show that in the learning environment described by the axioms there is a positive probability of constructing $\wp$ on each trial.

Let $C \subseteq V^* \times \cdots \times V^*$ (n times) be the set of encoded stimulus instances of the f-computable concept C—without loss of generality in this context I identify the concept with its set of instances, and let $\neg C$ be the complement of C. We take as a presentation distribution of stimuli, where $((\langle f_1 \rangle), \ldots, \langle f_n \rangle)$ is the encoding representation of a stimulus,

$$P((\langle f_1 \rangle, \ldots, \langle f_n \rangle) \in C) = P((\langle f_1 \rangle, \ldots, \langle f_n \rangle) \in \neg C) = \frac{1}{2}.$$

Moreover, we design the experiment to sample from C and $\neg C$ in the following geometric fashion. Let f_i be the coding in V^* of feature i of stimulus σ and let $|f_i|$ be the number of symbols in f_i. Then $\sum |f_i|$ is the total number of symbols used to encode σ. We use a geometric distribution for the total number of symbols, and a uniform distribution for selecting among those of the same total number of symbols. (In a completely formalized theory, these assumptions about probabilistic selection of presented stimuli would be part of the axioms, which I have restricted here just to the register learning model, and have not included axioms on stimulus presentation or reinforcement procedures in any detail.)

Suppose now that initially *LTM* is nonempty. If the program stored in *LTM* correctly computes f, we are done. If the program does not for some stimulus σ, then by the assumptions just stated there is a fixed positive probability that σ will be presented on every trial and hence with probability one asymptotically *LTM* will be cleared by virtue of Axiom Er2.

The probability of then constructing $\wp$ is positive on every trial. The detailed calculation is this. First, let $\wp$ have m lines. By Axiom P2(i), the probability of constructing a program of exactly m lines is equal to $c(1-c)^{m-1}$. If line i is not a jump instruction, then by Axiom P2(ii), the probability of line i being of the desired form is greater than some $\epsilon_1 > 0$. And if line i is a conditional jump instruction, where the jump is to line n_i , then also by Axiom P2(ii), the probability of line i being exactly line i of program $\wp$ is equal to $\epsilon_2^2(1-\epsilon_2)^{n_i-1}$ for some $\epsilon_2 > 0$.

So, independent of trial number, the finite product of these probabilities is positive on every trial. Explicitly, let $i_1, \ldots, i_{m_1}$ be the lines that are not jump instructions and let $j_1, \ldots, j_{m_2}$ be the lines that are, with $m = m_1 + m_2$. Then

$$\text{(I)} \qquad \text{Prob of } \wp > \epsilon_1^{m_1} \prod_{i=j_1}^{i=j_{m_2}} \epsilon_2^2(1-\epsilon_2)^{n_i-1} \cdot \ c(1-c)^{m-1} > 0.$$

From this inequality, we infer at once that asymptotically $\wp$ will be learned with probability one, which completes the proof, except to remark that to prove the constructed program characterizes a partial recursive function is straightforward.

Criticisms of the purely asymptotic character of this theorem are as appropriate as they were in criticisms of the perceptron convergence theorem (Theorem 6) or the language-learning theorem of Wexler and Culicover (Theorem 4). The next section addresses these problems.

Role of hierarchies and more determinate reinforcement. For the theory of register-model concept learning, as formulated in Definition 2, we cannot improve on inequality (I). Treating it as an equality it is evident that for programs $\wp$ of any length learning will be very slow, much slower than we observe in most human learning and even much animal learning.

Within the framework of the present theory, the only practical hope for learning to occur in a reasonable time is to organize learning into a hierarchy of relatively small tasks to be mastered. It might be thought that this conclusion could be avoided by making the reinforcement more informativé or determinate than what was assumed in Axioms Er1 and Er2 above. There is something correct and important about this view, and it can be supported by detailed computations on significant examples. On the other hand, there is also a question of interpretation. For the completely deterministic reinforcement used in the proof of Theorem 1, we could regard conditioning of each internal state of the finite automaton as a task—here task is defined by what gets reinforced, and in this view, the most fine-grained hierarchy is created by completely deterministic reinforcement.

It will be useful to end with application of the theory to a small, familiar task, to show that the theory can be brought down to earth and applied to data. Of course, in the present context I shall not try to be serious about actual parameter estimation. The task selected is that of 5-year-old children learning the concept of a triangle by recognizing triangles when presented with triangles, quadrilaterals and pentagons.

I make the following assumptions about the register model being used by the children. (It has the sort of simplifications necessary in such matters.)

(i) The language V_1 has a single element, α, which is used for counting.

(ii) There are two feature registers, #1 for number of segments and #2 for size, with α = small, $\alpha\alpha$ = medium and $\alpha\alpha\alpha$ = large.

(iii) The conditional jump is either to a previous line or to 0 (for a nonexistent line and stop).

(iv) To simplify formulation, computations are made directly on the feature registers rather than first copying their contents to a working register. (Characterizing copying from one register to another in terms of the three types of primitive instructions is straightforward.)

(v) *Rp* is the single response register.

(vi) Let a be the probability of selecting the delete instruction, b the probability for the jump instruction, and $1 - a - b$ the probability of the place or add instruction.

(vii) Let p be the probability of selecting feature register 1, and $1 - p$ that of selecting feature register 2 for reference in a line of program.

A simple correct program is:

1.	D(1)	Delete α from register 1.
2.	D(1)	Delete α from register 1.
3.	D(1)	Delete α from register 1.
4.	Copy(1,Rp)	Copy the contents of register 1 in the response register Rp.

All programs, in the short form used here, must end in copying the contents of a feature or working register to the response register. A response is then made. So the probability of lines 1-3 is: $p^3a^3c(1-c)^2$, where c is the parameter for the distribution of number of lines introduced in Axiom P2(i).

It is important to recognize that many different programs will produce the correct response, and so the probability of a correct response is considerably greater than $p^3a^3c(1-c)^2$. The complexity of a full analysis even for the simple experimental situation considered is much greater if the task is recognition of quadrilaterals rather than triangles. Still, under reasonable assumptions, the probabilities of the correct programs that are near the minimum length should dominate the theoretical computation of a correct response.

The learning setup defined axiomatically here is in terms of its scope comparable to the definition of partial recursive functions or the definition of register machines for computing such functions—namely, the definitions apply to each function considered individually. But for extended learning of a hierarchy of concepts, the structure must be enriched to draw upon concepts that have been previously learned in order to reach a practical rate of learning. Here is a very simple example to illustrate the point. Consider a disjunctive concept made up of n disjoint cases. Only one register is required, the alphabet V_1 is the set $\{\alpha, \beta\}$, and there is no jump instruction, but only the four instructions for deleting letters on the left or adding them on the right. Let the program be at most 10 lines for each case. Then assuming a uniform distribution on sampling of instructions and of the number of lines (1 to 10), the probability of each program of at most 10 lines can be directly computed. More importantly in the present instance, we can easily compute the possible number of programs: 4 of length 1, 16 of length 2, and in general 4^n of length n, with $1 \leq n \leq 10$, for a total of $(4^{11} - 4)/3$, which is approximately 4^{10}. If now at the second stage programs are put together using only original instructions and the n subroutines from individual cases, with programs of length at most $2n$ permitted, then there are $[(n+4)^{2n+1} - (n+4)]/(n+3)$ possible programs, which is approximately $(n+4)^{2n}$. On the other hand, if a single program is developed in one step with $10n$ lines, the number of possible programs is approximately 4^{10n}. Consider, for example, the case $n = 3$. Then 4^{30} is many orders of magnitude larger than $7^6 + 3(4^{10})$. The details of this example are not important, and I have not attempted to fix them sufficiently to determine in each of the two approaches the number of possible programs that are correct. Ordinarily in both the hierarchical and nonhierarchical approach this number would be a very small percentage of the total. The gain from the hierarchical approach is evident enough.

More generally, clever ways of dynamically changing the probability of using a previously defined concept, i.e., its recognition program, are critical to actual machine learning, for example, and sound hypotheses about such methods seem essential to any sophisticated study of human or animal learning of an extended hierarchy of concepts. Of equal importance is the introduction of forms of information feedback richer than the simple sort postulated in Definition 2, but the mathematical study of alternatives seems still to be in its infancy—only the extreme cases are relatively well understood. Much human learning depends upon verbal instruction and correction, but an approximately adequate theory of this process of feedback is as yet out of reach from a fundamental standpoint. Various gross simplifying assumptions, as used, for example, in protocol analysis, seem uneliminable at the present time. This is one measure of how much remains to be done.

29

ON DERIVING MODELS IN THE SOCIAL SCIENCES

There is a long tradition in the physical sciences of deriving from qualitative or quantitative empirical assumptions differential equations that govern, at least in approximation, a great variety of physical phenomena. The derivations of some of these equations, for example, the Navier-Stokes equations for hydrodynamical fluids, are among the most important conceptual analyses in the history of physics. The derivation of classical differential equations in the physical sciences is also of great importance in a workaday environment, where particular equations are derived to model particular circumstances, which in themselves may be of considerable practical importance, but are not of universal scientific interest. What is central and important to the applications of physical theories to empirical phenomena is the long and robust tradition of such derivations.

Corresponding methods are not as well developed for the social sciences, and it is evident enough by a glance at principal publications in the social sciences in which models are proposed and tested, that the derivation of differential equations does not dominate in any sense the derivation of models in the social sciences.

In this paper I examine five different kinds of derivation. The first are measurement representations, the second are regression-type models,

*Reprinted from *Journal of Mathematical & Computer Modelling*, **14** (1990), 21-28.

and the third are non-Markov observable models with process assumptions. The fourth are discrete Markov models with unobservable variables playing a theoretical role, and finally, the fifth exemplifies the classical derivation of a differential equation, so characteristic of the physical sciences.

At the end of the paper I will also discuss some of the general issues raised by these various methods of derivation. I touch at least briefly on questions of axiomatization, the use of theoretical or unobservable variables, and the relative importance of measurement procedures.

1. MEASUREMENT REPRESENTATIONS

In almost all scientific disciplines, it is understood that there can be neither precise control nor prediction of phenomena without measurement. In disciplines, especially in the social sciences, that do not have a long tradition of quantitative theory, the formulation of a theory of measurement can play a central role in scientific investigation. We can describe this general activity as a derivation of measurement models where, when it is done from a fundamental standpoint, it has the following character. Qualitative empirical relations and operations that can be handled in an explicit way experimentally form the basis of a qualitative axiomatic theory of measurement procedures. From an experimental or empirical standpoint the problem is to show to what extent the axioms postulated can actually be tested experimentally. From a formal standpoint on the other hand, the central task is to show that the axioms are formally adequate, that is, that a numerical measurement representation can be derived for any realization of the qualitative relations and functions that are the basis of the empirical axioms. The general subject has a long history and development which I shall not attempt to describe here. I shall concentrate rather on a simple but significant example, namely, the efforts to understand qualitative probability orderings. Both for reasons of psychological interest as reflected in the empirical study of beliefs and decisions, and for normative reasons connected with Bayesian approaches to statistics, the theory of qualitative probability relations has now a rather sophisticated development. To give a sense of how measurement representations are set up and studied I consider this one example, which sets aside any question of the measurement of utility and concentrates entirely on the measurement of subjective probability.

Let Ω be a nonempty set and let $\Im$ be an algebra of events on Ω, i.e., an algebra of sets on Ω. Let $\succeq$ be a qualitative ordering on $\Im$. The interpretation of $A \succeq B$ for two events A and B is that A is *at least as* *probable* as B. A (finitely additive) probability measure P on $\Im$ is *strictly agreeing* with the relation $\succeq$ if and only if, for any two events A and B in $\Im$,

$$P(A) \geq P(B) \text{ iff } A \succeq B.$$

A variety of conditions that guarantee the existence of a strictly agreeing measure is known. Without attempting a precise classification, the sets of conditions are of the following sorts: (i) sufficient but not necessary conditions for existence of a unique measure when the algebra of events is infinite (Koopman, 1940; Savage, 1954; Suppes, 1956); (ii) sufficient but not necessary conditions for uniqueness when the algebra of events is finite or infinite (Luce, 1967); sufficient but not necessary conditions for uniqueness when the algebra of events is finite (Suppes, 1969); (iv) necessary and sufficient conditions for existence of a not necessarily unique measure when the algebra of events is finite (Kraft, Pratt, & Seidenberg, 1959; Scott, 1964; Tversky, 1967). A rather detailed discussion of these various sets of conditions is to be found in Chapters 5 and 9 of Krantz, Luce, Suppes, and Tversky (1971).

A large literature which still continues shows that it is difficult to give simple necessary and sufficient qualitative conditions just in terms of events. On the other hand, simplification is relatively easy if we introduce some auxiliary concepts. In the present case the move is from an algebra of events to an algebra of extended indicator functions for the events. By this latter concept we mean the following. As before, let Ω be the set of possible outcomes and let $\Im$ be an algebra of events on Ω, i.e., $\Im$ is a nonempty family of subsets of Ω, and is closed under complementation and union, i.e., if A is in $\Im$, $\neg A$, the complement of A with respect to Ω, is in $\Im$, and if A and B are in $\Im$ then $A \cup B$ is in $\Im$. Let A^c be the indicator function (or characteristic function) of event A. This means that A^c is a function defined on Ω such that for any ω in Ω,

$$A^c(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A. \end{cases}$$

The algebra $\Im^*$ of *extended* indicator functions relative to $\Im$ is then just the smallest semigroup (under function addition) containing the indicator functions of all events in $\Im$. In other words, $\Im^*$ is the intersection of all sets with the property that if A is in $\Im$ then A^c is in $\Im^*$, and if A^* and B^* are in $\Im^*$, then $A^* + B^*$ is in $\Im^*$. It is easy to show that any function A^* in $\Im^*$ is an integer-valued function defined on Ω. It is the extension from indicator functions to integer-valued functions that justifies calling the elements of $\Im^*$ extended indicator functions.

The qualitative probability ordering must be extended from $\Im$ to $\Im^*$, and the intuitive justification of this extension must be considered. Let

A^* and B^* be two extended indicator functions in $\Im^*$. Then, to have $A^* \succeq B^*$ is to have the expected value of A^* equal to or greater than the expected value of B^*. As should be clear, extended indicator functions are just random variables of a restricted sort. The qualitative comparison is now not one about the probable occurrences of events, but about the expected value of certain restricted random variables. The indicator functions themselves form, of course, a still more restricted class of random variables, but qualitative comparison of their expected values is conceptually identical to qualitative comparison of the probable occurrences of events.

The axioms are embodied in the definition of a qualitative algebra of extended indicator functions. Several points of notation need to be noted. First, Ω^c and $\emptyset^c$ are the indicator or characteristic functions of the set Ω of possible outcomes and the empty set $\emptyset$, respectively. Second, the notation nA^* for a function in $\Im^*$ is just the standard notation for the (functional) sum of A^* with itself n times. Third, the same notation is used for the ordering relation $\Im$ and $\Im^*$, because the one on $\Im^*$ is an extension of the one on $\Im$: for A and B in $\Im$,

$$A \succeq B \text{ iff } A^c \succeq B^c.$$

Finally, the strict ordering relation $\succ$ is defined in the usual way: $A^* \succ B^*$ iff $A^* \succeq B^*$ and not $B^* \succeq A^*$.

DEFINITION. *Let Ω be a nonempty set, let $\Im$ be an algebra of sets on Ω, and let $\succeq$ be a binary relation on $\Im^*$, the algebra of extended indicator functions relative to $\Im$. Then the qualitative algebra $(\Omega, \Im^*, \succeq)$ is* qualitatively satisfactory *if and only if the following axioms are satisfied for every A^*, B^*,* and *C^* in $\Im^*$:*

1. *The relation $\succeq$ is a weak ordering of $\Im^*$;*
2. $\Omega^c \succ \emptyset^c$;
3. $A^* \succeq \emptyset^c$;
4. $A^* \succeq B^*$ *iff* $A^* + C^* \succeq B^* + C^*$;
5. *If $A^* \succ B^*$ then for every C^* and D^* in $\Im^*$ there is a positive integer n such that*

$$nA^* + C^* \succeq nB^* + D^*.$$

These axioms should seem familiar from the literature on qualitative probability. Note that Axiom 4 is the additivity axiom that closely resembles de Finetti's additivity axiom for events: *If $A \cap C = B \cap C = \emptyset$, then $A \succeq B$ iff $A \cup C \succeq B \cup C$.* As we move from events to extended indicator functions, functional addition replaces union of sets. What is formally of importance about this move is seen already in the exact formulation of Axiom 4. The additivity of the extended indicator functions is unconditional—there is no restriction corresponding to $A \cap C = B \cap C = \emptyset$. The absence of this restriction has far-reaching formal consequences in permitting us to apply without any real modification the general theory of extensive measurement.

THEOREM. *Let Ω be a nonempty set, let $\Im$ be an algebra of sets on Ω, and let $\succeq$ be a binary relation on $\Im$. Then a necessary and sufficient condition that there exist a strictly agreeing probability measure on $\Im$ is that there is an extension of $\succeq$ from $\Im$* to *$\Im^*$ such that the qualitative algebra of extended indicator functions $(\Omega, \Im^*, \succeq)$ is qualitatively satisfactory. Moreover, the expectation function on $\Im^*$ is unique.*

There is not room to give the proof here, which is given in Suppes and Zanotti (1976) where much of the preceding discussion also is formulated. The important thing about the theorem is that its proof uses in a straightforward way standard results in the general theory of extensive measurement. Extension to the case of conditional probability is to be found in Suppes and Zanotti (1982).

Once the theorem is available we then have reduced the qualitative probability ordering of events to the familiar theory of what are often called elementary or simple random variables.

2. REGRESSION-TYPE MODELS

To illustrate the ideas, I examine an analysis of the causes of the levels of consumption in terms of individual disposable income in a given population. In this kind of analysis, typical of regression analysis, there is no claim that the amount of disposable income is the only cause of the level of consumption, but almost everyone would expect it to be a principal cause. The interesting question is whether present consumption is influenced by past disposable income.

To examine some models, let

c_{it} = consumption in time period t, by individual or household i
d_{it} = disposable income of i in period t,

and for aggregation of n individuals, let

$$c_t = \sum_{i=1}^{n} c_{it}$$

and

$$D_t = \sum_{i=1}^{n} d_{it}.$$

(For a good discussion of such aggregation and other methodological aspects of the models considered in this section, see Malinvaud (1966, Chapter 4).)

In terms of the notation defined, an obvious linear model is

$$C_t = a_0 \sum_{\tau=1}^{T} b(t-\tau) D_{t-\tau} + e + \epsilon_t \tag{1}$$

where a_0 and e are constants, $b(t-\tau)$ is a constant for period $t-\tau$ and ϵ_t is the error term for period t. In the usual probabilistic formulation it is assumed that the expectation of ϵ_t is zero.

A classic study by Friedman (1957) of annual data on consumption and disposable income for heads of household in the United States for the period 1905-1951 but excluding the war years yields the following numerical version of (1):

$$\begin{aligned} C_t = \;& 0.29D_t + 0.19D_{t-1} + 0.13D_{t-2} + 0.09D_{t-3} \\ & +0.06D_{t-4} + 0.04D_{t-5} - 4 \end{aligned} \tag{2}$$

The point in the present context is that much empirical economics naturally uses as models the obvious sort of linear regression model exemplified by equations (1) and (2). It is almost never claimed that such models are an exact account of the phenomena in question. The point is to sustain the argument that a fair amount of the variation in the phenomena is accounted for by such a linear regression model.

The right way to think about regression models in my view is that they are the beginning point of an analysis, not the end. On the other hand, they are often successful in accounting for a high percentage of the variation in the phenomena and therefore the search for more detailed models is for many purposes not as important as it might appear to be.

3. NON-MARKOVIAN OBSERVABLE MODELS WITH PROCESS ASSUMPTIONS

The linear learning model that I actually use as an example here is one that has been widely used in mathematical psychology and also in the development of control processes in engineering. I will describe the model from a conceptual standpoint in terms of the learning of an organism. Also it is assumed that the situation is one of discrete trials rather than continuous time but the generalization to continuous time is also of interest.

For simplicity, let us assume that on every trial the organism can make exactly one of two responses, A_1 or A_2, and after each response it receives a reinforcement, E_1 or E_2 of one of the two possible responses. A learning parameter θ, which is a real number such that $0 < \theta \leq 1$, describes the rate of learning in a manner to be made definite in a moment. A possible realization of the theory is an ordered triple $\mathcal{X} = (X, P, \theta)$ of the following sort. X is the set of all sequences or ordered pairs such that the first member of each pair is an element of some set A and the second member an element of some set B, where A and B each have two elements. Intuitively, the set A represents the two possible responses and the set B the two possible reinforcements. P is a probability measure on the σ-algebra of cylinder sets of X, and θ is a real number as already described. To define the models of the theory, we need a certain amount of notation. Let $A_{i,n}$ be the event of response A_i on trial n; $E_{j,n}$ the event of reinforcement E_j on trial n, where $i, j = 1, 2$; and for x in X let x_n be the equivalence class of all sequences in X which are identical with x through trial n. We may then characterize the theory by the following set-theoretical definition.

A triple $\mathcal{X} = (\mathcal{X}, \mathcal{P}, \theta)$ *is a* linear learning model *if and only if the following two axioms are satisfied:*

A1 *If* $P(E_{i,n}A_{i',n}x_{n-1}) > 0$ *then*
$P(A_{i,n+1}|E_{i,n}A_{i',n}x_{n-1}) = (1-\theta)P(A_{i,n}|x_{n-1}) + \theta.$
A2 *If* $P(E_{j,n}A_{i',n}x_{n-1}) > 0$ *and* $i \neq j$ *then*
$P(A_{i,n+1}|E_{j,n}A_{i',n}x_{n-1}) = (1-\theta)P(A_{i,n}|x_{n-1}).$

As is clear from the two axioms, this linear response theory is intuitively very simple. The first axiom just says that when a response is reinforced, the probability of making that response on the next trial is increased by a simple linear transformation. The second axiom says that if some other response is reinforced, the probability of making the response is decreased by a second linear transformation. These linear learning models are examples of *chains of infinite order*, so called because the dependence on the past does not terminate after some fixed time or fixed number of trials. Still, we find it very natural for a variety of reasons to think in terms of Markov processes, with the present state absorbing all needed information about the past, and so a definite effort has been made to redefine the concept of state for such chains of infinite order as linear learning models. When the reinforcement occurring on trial n is probabilistically dependent at most on the immediately preceding response on that trial, then the response probabilities can be taken as the states, and it is easy to show—under the restriction stated—that the process is Markov. Extensive examples of this Markovian approach are developed, as well as the general theory, in Norman (1972). However, even for the following simple Markov reinforcement schedule, this approach will not work.

Let $P(E_{1,1}) = \gamma$, and

$$\begin{array}{c|cc} n \diagdown n+1 & E_1 & E_2 \\ \hline E_1 & \alpha & 1-\alpha \\ E_2 & 1-\beta & \beta \end{array}$$

Then the Markovian approach of Norman will not work, as is easy to show. To get to a Markov model we must enlarge the set of psychological concepts. This kind of model is discussed in the next section. The properties of this linear learning model are studied extensively in Estes and Suppes (1959a). The model is generalized to a continuum of responses in Suppes (1959).

Here is a simple asymptotic result taken from Estes and Suppes (1959a) to illustrate the sort of testable quantities that are derived for a given schedule of reinforcement or feedback. Let $\pi_{j,k}(\nu)$ be the probability that reinforcing event E_k will occur on trial n given that response A_j occurred ν trials earlier. Then

$$\lim_{n\to\infty} P(A_{1,n}) = \frac{\pi_{21}(\nu)}{1 - \pi_{11}(\nu) + \pi_{21}(\nu)}.$$

4. MARKOV MODELS WITH UNOBSERVABLE THEORETICAL VARIABLES

The fundamental theory of conditioning I develop briefly is a variant of stimulus-sampling theory first formulated by Estes (1950). The axioms are stated formally in Estes and Suppes (1959b, 1974). Here I give only an informal statement.

Conditioning axioms.

C1. *On every trial each stimulus element is conditioned to at most one response.*

C2. *If a stimulus element is sampled on a trial, it becomes conditioned with probability c to the response (if any) that is reinforced on that trial; if it is already conditioned to that response, it remains so.*

C3. *If no reinforcement occurs on a trial, there is no change in conditioning on that trial*

C4. *Stimulus elements that are not sampled on a given trial do not change their conditioning on that trial.*

C5. *The probability c that a sampled stimulus element will be conditioned to a reinforced response is independent of the trial number and the preceding pattern of events.*

Sampling axioms.

S1. *Each available stimulus element is sampled with a probability that is independent of the sampling of any other available element.*

S2. *Given the set of stimulus elements available for sampling on a trial, the probability of sampling a given element is independent of the trial number and the preceding pattern of events.*

Response axioms.

R1. *The probability of a given response is the proportion of sampled stimulus elements conditioned to that response, given that at least one such conditioned element is sampled.*

R2. *If no sample stimulus element is conditioned, then there is a probability p_i that response i will occur.*

R3. *The guessing probability p_i of response i, when the sampled stimulus element is not conditioned, is independent of the trial number and the preceding pattern of events.*

There are two important remarks in the present context to make about these axioms. Particular models that are derived from these axioms have two essential features. First is the Markov character of the sequence of state of conditioning random variables. We can represent the state of conditioning by an appropriate random variable or in many instances, we can do this by a random variable that simply represents the number of stimuli conditioned to each response. In either case these random variables will satisfy Markov assumptions which can be rigorously derived from the explicit formulation of the theory.

The second feature is that the observation of the actual stimulus elements sampled or the observation of the conditioning relations that obtain is not possible. Both of these concepts are theoretical in the ordinary applications of the theory, and play throughout the analysis of experiments a theoretical role. This is not to say that in a more enlarged theory and in a more physiological setting one might attempt to make direct observations. This is to some extent the main business of many neurophysiologists. But as such a theory is ordinarily applied in psychology these concepts do not represent observable variables.

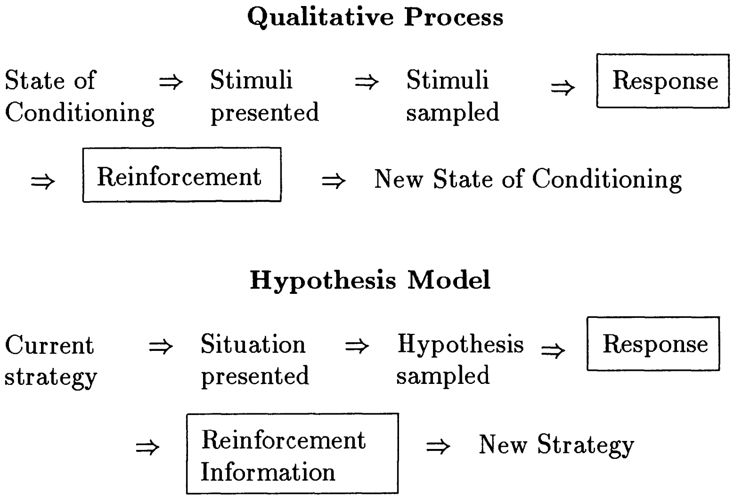

Figure 1. Similar structure of conditioning and hypothesis models of learning

Another important remark of a different sort is that the kind of theory just formulated in terms of conditioning—a language especially sympathetic to current work in neurophysiology, can be replaced by language more sympathetic to the interests and concepts of many cognitive scientists. In this case we replace the concept of conditioning by that of a strategy where a strategy consists of a class of hypotheses. I shall not work out that correspondence here but it is a familiar one and the real question is which kind of language a given scientist finds most suggestive for use in experimental investigations. Without spelling things out explicitly, the parallel analysis in Figure 1 of the conditioning model and of the hypothesis model shows how congruent they are in representing the same sequence of events. Of course we can no more observe the hypothesis being used by a subject than we can observe the state of conditioning. (It is also important to recognize that subjects' abilities to verbalize strategies of any complexity or subtlety are rather limited.)

The general theorem that is proved in Estes and Suppes (1974) is this.

THEOREM. *If a stimulus sampling model has an experimenter's schedule of reinforcement that is dependent only on a fixed finite number of preceding trials, then there is a sequence of random variables combining the reinforcement schedule and state of conditioning such that this sequence of random variables is a finite-state Markov chain.*

(In stating the theorem I have omitted the technical notation which has a rather direct intuitive meaning. The reader is referred to Estes and Suppes (1974) for details.)

5. CONTINUOUS MODELS

I mentioned at the beginning the importance of the derivation of differential equations in the physical sciences and in a wide range of applications in engineering. Such derivations of differential equations also play a role in the social sciences, more in economics than elsewhere, but there are a variety of examples of differential equations, and their derivation from empirical assumptions or principles, in psychology and sociology. I want to give one example here from my own work in educational psychology. The example also exhibits several features characteristic of the use of differential equations in the physical sciences. For example, the underlying phenomena are discrete but for purposes of analysis it is reasonable to derive a differential equation for the continuous approximation.

Underlying the particular application considered here is the desire to be able to predict student progress in a course, particularly a course that is computer-based, and in addition individualized. Classic evaluation of a new curriculum in the schools has an important unsatisfactory feature. This is the evaluation by comparing pretests and posttests with an analysis of post-test grade placement distributions as a function of pretest distribution and exposure to the new curriculum. The unsatisfactory part is the wait-and-see approach required. In contrast, it is a characteristic feature of the physical sciences to continually seek for models that predict phenomena and that can be used for control purposes as in engineering. It is such an approach to analysis of student progress in a curriculum that I want to consider here. I summarize research that has taken place over many years beginning with Suppes, Fletcher, and Zanotti (1976).

Let us assume as already indicated that the student is progressing individually through a course, and let $I(t)$ be the total information presented to the student up to time t. I say here *total information* but we could also give a formulation in terms of *skills* and think of the development procedurally rather than declaratively. Let $y(t)$ be the student's course position at time t. Note that for simplification of notation I have omitted a subscript for a particular student. It is understood that the notation used here applies only to an individual student not to averages of students. The stochastic averaging involved is averaging over the variety of skills or information presented to the student to refer to his mean position and mean information. I shall not make these stochastic assumptions explicit any further, because only the mean theory for the individual student will be developed here.

The first assumption is that the process is additive using for simplicity of concept a discrete variable n. We may write the additivity assumption as follows:

$$\text{Additive: } I(n) - I(n-1) = \alpha, \tag{1}$$

which we can then express in terms of a derivative as in (2)

$$\frac{\dot{I}(n)}{I(n)} = \frac{\alpha}{n\alpha} = \frac{1}{n}, \tag{2}$$

with the proper expression for continuous time shown as equation (3)

$$\frac{\dot{I}(t)}{I(t)} = \frac{1}{t}. \tag{3}$$

We then make the second strong assumption that the position in the course is proportional to the information introduced, that is

$$y(t) \approx I(t). \tag{4}$$

Combining (3) and (4) we have equation (5)

$$\frac{\dot{y}(t)}{y(t)} \approx \frac{1}{t}, \tag{5}$$

which we then integrate to obtain equation (6)

$$ln\ y(t) = k\ ln\ t + ln\ b, \tag{6}$$

which we may express as equation (7), so that $y(t)$ is a power function of t:

$$y(t) = bt^k. \tag{7}$$

We have extensively applied this power function to the analysis of student data from courses in elementary mathematics and reading that are computer based. This analysis has been going on for a number of years at Computer Curriculum Corporation. I show here in Figure 2 the data for a student in elementary mathematics and the theoretical power function where only the coefficient b and the exponent k are estimated, and similarly for another student as shown in Figure 3 for an elementary course in reading. The grade placement shown on the ordinate of each figure gives a sense of the achievement level of the student. On the abscissa is shown the amount of time spent at the computer by the student. An interesting fact about these curves is that they extend over a very

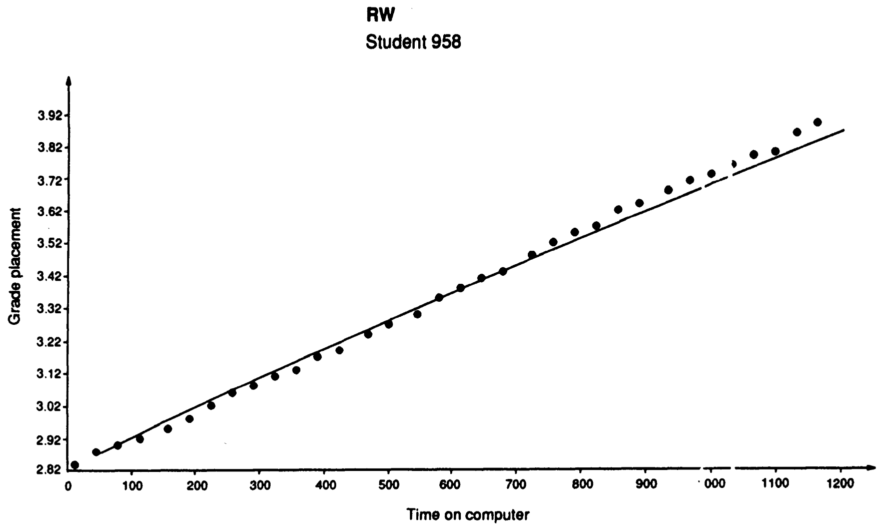

Figure 2. Comparison of observed and theoretical individual student trajectory in computer-based elementary-school mathematics course.

considerable amount of time both in terms of computer time but also in terms of calendar time, which is for a major part of the school year.

We have gone on to use this analysis to estimate individual student trajectories on the basis of the first several months of the school year and to predict what the students will gain in grade placement in a given subject during the school year. On the basis of this prediction intervention can be made as necessary. I shall not try to describe the character of that intervention here, but just emphasize it is important to pass from data analysis to prediction in order to take full advantage of the opportunities, and to fulfill the original intention for which the theory was developed, namely to give a predictive-control feature to the introduction of a new curriculum.

6. SOME GENERAL ISSUES

Axiomatics. It has been remarked on a variety of occasions that use of axiomatic methods is more common in the social sciences than in the

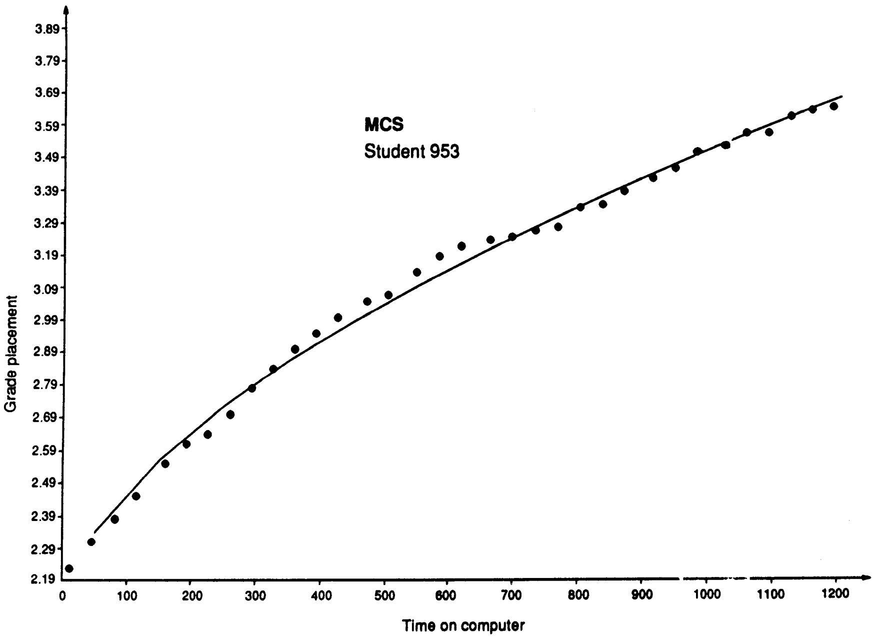

Figure 3. Comparison of observed and theoretical individual student trajectory in computer-based elementary-school reading course.

physical sciences. This remark has usually been mildly pejorative in spirit. What I have tried to emphasize here is the derivation and formulation of a variety of models rather than the pure use of axiomatic methods as reflected in the explicit and precise statement of general theories. The example I gave for measurement exemplifies of course the axiomatic method in pure form. In contrast, the last example, in which a power function is derived for predicting student progress in a course, is not really axiomatic in form but is in the spirit of classical derivations of differential equations where some particular assumptions are made that form the basis for the derivation. The point here is that the empirical assumptions from which the differential equations are derived are not elevated to the status of axioms of a general theory. The distinction I have in mind is not one that is important to make completely precise but it is familiar enough in the physical sciences. Presented with a new physical situation, as for example the motion of a robotic arm, it is important to derive the equations of motion from the given physical constraints, but we do not ordinarily think of the statement of the physical assumptions peculiar to the particular robot arm being studied as having the status of axioms. I think the same thing is true of several of the models I have discussed here. We would not want to think of the kind of derivations I have talked about as exemplifying the axiomatic method in any pure form. Above all, there is really no process of actual derivation from axioms in the case of the linear regression model discussed and this is characteristic of many applications of linear regression though by no means all.

It is also important to recognize there is a long tradition of beginning with something fairly close to the surface, fairly phenomenological in spirit, and then later attempting to derive the phenomenological equations from some more fundamental assumptions. This is done in Estes and Suppes (1959b) for the linear learning models discussed above by assuming that the sampled stimuli asymptotically approach infinity with some restrictive assumptions on how this approach takes place. The linear learning model can then be derived in the limit from the stimulus sampling model. But this kind of foundational investigation, important as it is, is not the focus here and is not as important probably for most scientific purposes as having available a variety of methods for deriving models that can be fairly directly tested.

Use of unobservable theoretical variables. In contrast to the last remark about axiomatics, I also want to emphasize the importance I attach to getting theory below the phenomenological surface. A long history of developments in physics indicates the importance of this. The same would seem to be true of models in the social sciences. In terms of the kind of examples I have discussed drawn from learning and cognition, it is evident that our scientific thinking will go beyond just observable data almost naturally. It also seems very likely that the internal mechanisms and the associated theoretical concepts that characterize the mechanisms will be unobservable in detail for the indefinite future if not forever in the case of most human learning and cognition. Any program for the elimination of unobservable theoretical variables would seem to me to be completely mistaken, and contrary to almost all the deeper developments in present-day scientific psychology, not to speak of the important role of assumptions about the unverified and often unverifiable assumptions about individual choices and preferences in neo-classical economic theory.

Importance of measurement. On the other hand, it is important to stress the importance of measurement in the social sciences where procedures do not have the long history of development they have had in the physical sciences. I gave at the beginning just one example, the measurement of subjective degree of belief, or subjective probability. But in all of the social sciences problems of measurement are present and the theory needs continual development. It is also important to realize that often these developments are not simply in terms of phenomenological variables but are in terms of theoretical concepts as well. Perhaps the most salient one is the concept of utility in decision making, which is widely used in economics and psychology. The axioms of utility, or of expected utility, as usually stated, are not directly observable and the concept of an individual's utility function must be assigned a theoretical status. This does not mean that the measurement of this theoretical concept is not of great importance, for from the theory of its measurement many observable consequences can be derived.

Just because of the relatively recent development of methods of measurement in the social sciences, the theory of measurement continues to be of great importance and will undoubtedly continue to flourish, and especially, be the source of the derivation of many models in the social sciences. On the other hand, I would not want to give a sense that the derivation of models in the social sciences can in any sense be reduced simply to problems of measurement. Theories of measurement are important, but they are in no sense the whole story.

What I have tried to give a sense of in this paper is the plurality of methods that should be used and should be available to the social scientist deriving models. It would seem to me to be a piece of folly to argue that one method is better than another, for the contexts of use are sufficiently varied and rich in differentiation.

30

THE PRINCIPLE OF INVARIANCE WITH SPECIAL REFERENCE TO PERCEPTION

The principle of invariance is now a familiar one in psychology, especially because of its prominent role in the theory of measurement. The relation of invariance to the meaningfulness of various statistics for various scales of measurement is a particularly salient example that has received much discussion in the literature for over thirty years. Perhaps the most notorious case is that of whether standard intelligence tests have more than ordinal properties.

The concept of invariance has an older history in geometry. It came to prominence in the nineteenth century with the work of Felix Klein and his view that the transformations themselves could be taken as primitive geometrical concepts. For example, Euclidean geometrical properties are characterized by their invariance under the group of Euclidean motions. There is, however, a lot more to geometry than Kleinian-type transformations, but the principle of invariance remains an important one.

The principle has also achieved importance in physics, and in many ways has had a more prominent role in the twentieth century in physics than in geometry. Both invariance under the Galilean transformations of classical physics, the Lorentz transformations of special relativity, and the

*Reprinted from *Mathematical psychology: Current developments*, J. Doignon and J. Falmagne (Eds.), 1991, pp. 35–53. New York: Springer-Verlag.

transformations of general relativity have been central ideas in modern physics. But the central role of invariance does not stop there. Of almost equal importance is the relation between a given invariance or symmetry property of a physical system and a corresponding conservation law. Such relations were first studied by Emmy Noether, and now many Noetherian theorems play an important role in physical theories, for example, theorems on the conservation of energy, momentum and angular momentum in both classical and quantum mechanics.

My purpose in this lecture is to examine the role of the principle of invariance in perception, which moves away from the concerns of invariance in measurement, as do the Noetherian theorems in physics. On the other hand, I shall not attempt anything like a systematic survey, for the ways in which concepts of invariance enter into perception are many and varied. Rather, I shall focus on two main areas, both concerned primarily with visual perception. The first is the kind of invariance related to ordinary experience and the use of ordinary language. Here I shall be especially concerned with the relation between perception and geometry, less so with the ways in which physical concepts also influence, or are part of the meaning of, ordinary perceptual descriptions. In the second part I examine in some detail whether anything like the invariance characteristic of the physical concepts of space or space and time, as reflected in Galilean or Lorentz invariance, can be expected to hold for the perception of visual space. And if not, are there other, perhaps more limited, principles of symmetry that are salient.

A preliminary remark is needed about the way the principle of invariance enters the analyses I am concerned with. In most of the literature in psychology or physics on invariance, a given theory is held fixed and unchanged, and the invariance of some relation, function or proposition with respect to the theory is examined. For example, in a purely ordinal theory of measurement the mean of a set of numerical data is not invariant with respect to the theory, for arbitrary monotone transformations of the numerical data are permitted by a purely ordinal theory, and the mean is obviously not invariant under an arbitrary monotone transformation. Here I pursue a different path. In Part I various ordinary spatial expressions are introduced with their ordinary meaning, and the question asked is this. What is the natural geometrical theory with respect to which each expression is invariant? In Part II, similar questions are asked about the results of various experiments on visual space which challenge the thesis that visual space is Euclidean. As already hinted at, I also shall follow the practice in geometry and physics of relating invariance under a given group of transformations to a principle of symmetry. Invariance implies symmetry and symmetry implies invariance.

1. GEOMETRY OF QUALITATIVE VISUAL PERCEPTIONS AS EXPRESSED IN ORDINARY LANGUAGE

The concepts, results and problems that fall under the general heading of this part of the lecture have not been studied very intensely in the past but have received more attention in the last ten years or so. I mention in this connection especially Bowerman (1989), Crangle and Suppes (1989a), and Levelt (1982, 1984). The references in these publications provide good leads back into the older literature. A typical problem that I have in mind here is the sort of thing that Levelt has investigated thoroughly. For example, the way language is used to give spatial directions, and the limitations of our ability to give such directions or describe visual scenes with accuracy. In the case of my own earlier work with Crangle, we were concerned especially to analyze the geometry underlying the use of various common prepositions in English.

The results of that analysis can be easily summarized in Table 1 which is reproduced with some modifications from Crangle and Suppes (1989a). The kinds of geometry referred to are standard, with the exception perhaps of the geometry of oriented physical space. For example, in the case of the sentence *The pencil is in the box* (where it is assumed the box is closed), it is clear that only a purely topological notion of invariance is needed. On the other hand, in the sentence *Mary is sitting between Jose and Maria*, it is easy to see that the geometry needs to be affine. And once the idea of a metric is introduced, as in the sentence *The pencil is near the box*, we must go from affine geometry to some underlying notion of congruence as reflected in Euclidean geometry. Although we may be more refined in the analysis, note that in this sentence it is quite satisfactory to use absolute geometry which is Euclidean geometry minus the standard Euclidean axiom. This axiom asserts that given a point a and a line L on which the point does not lie, then there exists at most one line through a in the plane formed by aL which does not meet the line. We get hyperbolic rather than Euclidean geometry by adding the negation of this axiom to absolute geometry. It seems to me that the notion of nearness used in ordinary talk is satisfied well enough by the congruence relation of absolute geometry—a still weaker geometry of rigid bodies is considered later. In fact, relatively technical geometrical results are required to move us from absolute to Euclidean geometry, the sort of technical facts required, for example, in architecture and construction. Another way of noting the adequacy of absolute geometry to express many of the elementary results of Euclidean geometry is that the first 26 propositions of Book I of Euclid's *Elements* are provable in absolute geometry. On the other hand, in the case of the preposition *on*, it is clear that a notion

Topology	*The pencil is in the box.* (*box closed*) *One piece of rope goes over and under the other.*
Affine geometry	*The pencil is in the box.* (*box open*) *Mary is sitting between Jose and Maria.*
The geometry of oriented physical space	*The book is on the table.* *Adjust the lamp over the table.*
Projective geometry	*The post office is over the hill.* *The cup is to the left of the plate.*
Geometries that include figures and shapes with orienting axes	*The dog is in front of the house.* *The pencil is behind the chair.*
Geometry of classical space-time	*She peeled apples in the kitchen.*

Table 30.1. Kinds of geometry and examples of prepositional use.

of vertical orientation is required, a notion completely absent from Euclidean geometry, and in fact not definable within Euclidean geometry. A different kind of orientation is required in the case of objects that have a natural intrinsic orientation. Consider for instance, the sentence given in Table 1, *The dog is in front of the house.* Finally, in the case of many processes it is not sufficient to talk about static spatial geometry but for a full discussion one needs the assumption of space-time. An example is given at the end of Table 1.

What I now want to look at are the kinds of axioms needed to deal with the cases of geometry that are not standard. It would not be appropriate simply to repeat standard axioms for topological, projective, affine, absolute, and Euclidean geometry. A rather thorough recent presentation of these geometries is to be found in Suppes, et. al., (1989, Ch. 13). What I shall do is make reference to the primitive notions on which these various axioms are based as given in the above reference.

Oriented physical space. Undoubtedly, the aspect of absolute or Euclidean geometry which most obviously does not satisfy ordinary talk about spatial relations is that there is no concept of vertical orientation. Moreover, on the basis of well-known results of Tarski concerning the fact that no nontrivial binary relations can be defined in Euclidean geometry, the concept is not even definable in Euclidean geometry. For definiteness I have in mind as the primitive concepts of Euclidean geometry the affine ternary relation of betweenness for points on a line and the concept of congruence for line segments. In many ways I should mention however, it is more convenient to use the notion of parallelism and perpendicularity, and in any case I shall assume these latter two notions are defined. There are many different ways, of course, of axiomatizing as an extension of Euclidean geometry the concept of verticality. One simple approach is to add as a primitive the set V of vertical lines, and then to add axioms of the following sort to three-dimensional Euclidean geometry. *Given any point a there is a vertical line through a. If K is a vertical line and L is parallel to K, then L is a vertical line.*

There are however, difficulties with this approach of two different kinds. The first and conceptually the most fundamental is that our natural notion of oriented physical space as we move around in our ordinary environment is that the orientation of the space, both vertically and horizontally, is fixed uniquely. We do not have arbitrary directional transformations of verticality nor of horizontal orientation. We naturally have a notion of north, east, south, and west, with corresponding degrees. Secondly, and closely related to this, very early in the discussion of the nature of physical space, it was recognized that we have difficulties with treating the surface of the earth as a plane with horizontal lines being in this plane and vertical lines being perpendicular to them. The natural geometry of oriented physical space in terms of ordinary experience—a point to be expanded upon in a moment—is in terms of spherical geometry, with the center of the earth being an important concept, as it was for Aristotle and Ptolemy in ancient times, who in many ways most clearly expressed ideas about the nature of physical space.

Aristotle directly uses perceptual evidence as part of his argument for the conclusion that the earth is spherical *On the heavens*, Book II, Ch. 14, 297b): *if the earth were not spherical, eclipses of the moon would not exhibit the shapes they do, and observation of the stars would not show the variation they do as we move to the north or south.*

Ptolemy's argument in the *Almagest* is even better and because it will not be familiar to many readers, I quote in full, for the arguments are perceptual throughout. This passage occurs in Book I, Section 4 of the *Almagest*, written more than four hundred years later than Aristotle's work.

> That the earth, too, taken as a whole, is sensibly spherical can best be grasped from the following considerations. We can see, again, that the sun, moon and other stars do not rise

and set simultaneously for everyone on earth, but do so earlier for those more towards the east, later for those towards the west. For we find that the phenomena at eclipses, especially lunar eclipses, which take place at the same time [for all observers], are nevertheless not recorded as occurring at the same hour (that is at an equal distance from noon) by all observers. Rather, the hour recorded by the more easterly observers is always later than that recorded by the more westerly. We find that the differences in the hour are proportional to the distances between the places [of observation]. Hence one can reasonably conclude that the earth's surface is spherical, because its evenly curving surface (for so it is when considered as a whole) cuts off [the heavenly bodies] for each set of observers in turn in a regular fashion.

If the earth's shape were any other, this would not happen, as one can see from the following arguments. If it were concave, the stars would be seen rising first by those more towards the west; if it were a plane, they would rise and set simultaneously for everyone on earth; if it were triangular or square or any other polygonal shape, by a similar argument, they would rise and set simultaneously for all those living on the same plane surface. Yet it is apparent that nothing like this takes place. Nor could it be cylindrical, with the curved surface in the east-west direction, and the flat sides towards the poles of the universe, which some might suppose more plausible. This is clear from the following: for those living on the curved surface none of the stars would be ever-visible, but either all stars would rise and set for all observers, or the same stars, for an equal [celestial] distance from each of the poles, would always be invisible for all observers. In fact, the further we travel toward the north, the more of the southern stars disappear and the more of the northern stars appear. Hence it is clear that here too the curvature of the earth cuts off [the heavenly bodies] in a regular fashion in a north-south direction, and proves the sphericity [of the earth] in all directions.

There is the further consideration that if we sail towards mountains or elevated places from and to any direction whatever, they are observed to increase gradually in size as if rising up from the sea itself in which they had previously been submerged: this is due to the curvature of the surface of the water. (pp. 40–41)

Aristotle's concept of natural motion, which means that heavy bodies fall toward the center of the earth, is well argued for again by Ptolemy in Section 7 of Book I.

> ...the direction and path of the motion (I mean the proper, [natural] motion) of all bodies possessing weight is always and everywhere at right angles to the rigid plane drawn tangent to the point of impact. It is clear from this fact that, if [these falling objects] were not arrested by the surface of the earth, they would certainly reach the center of the earth itself, since the straight line to the center is also always at right angles to the plane tangent to the sphere at the point of intersection [of that radius] and the tangent. (pp. 43-44)

What is important to notice about this argument and similar but less clear arguments of Aristotle is that the notion of vertical or up is along radii extended beyond the surface of the earth, and not in terms of lines perpendicular to one given horizontal plane. Thus the proper perceptual notion of verticality is in terms of a line segment that passes through the center of the earth. The notion of horizontal is that of a plane perpendicular to a vertical line at the surface of the earth. What is also important here is that the strongest argument for this viewpoint is the perceptual evidence of the nature of natural falling motion of heavy bodies.

Ptolemy also uses observations of the motion of the stars and the planets to fix the direction of east and west, and also the poles of north and south. We use in ordinary experience such arguments as the rising and setting of the sun to fix the direction of east and west, and in a similar vein we fix the north and south poles.

Looked at from the standpoint of the symmetries or invariance of the standard group of Euclidean motions, these perceptual arguments about physical space—including of course perceptual arguments about gravity—reduce the group of Euclidean motions to the trivial automorphism of identity. This means that from a global standpoint the concept of invariance is not of any real significance in considering the perceptual aspects of oriented physical space. On the other hand the strong sense of the concept of global that is being used here must be emphasized. From the standpoint of the way the term *global* is used in general, there remain many symmetries of a less sweeping sort that are important and that are continually used in perceptual or physical analysis of actual space. Indeed this is a very intuitive outcome from the standpoint of ordinary perception. Combining both our visual sense of space and our sense of space arising from gravitational effects, it is wholly unnatural to think of anything like the group of Euclidean motions being the fundamental group for physical space in the sense of direct experience or perception. It is in fact a remarkable abstraction of the Greeks that orientation was not made a part of the original axioms of Euclid.

On the other hand, the notion of invariance in perception arises continually in a less global fashion in considering the symmetry of perceived figures or perceived phenomena arising from not just visual but also auditory or haptic data. A thorough development of symmetry in regular figures, especially two-dimensional ones, is given in Toth (1964). This is a subject with a rich history in both mathematics and arts, but I turn aside from it here to consider from a similar but different viewpoint the concept of invariance for what is expressed by a spatial preposition.

The classification of geometries for such prepositions given in Table 1 may be regarded as an *external* view from the standpoint of space as a whole. Perceptually, however, this is not the way we deal with the matter. In ordinary experience, we begin with the framework of a fixed oriented physical space as described above. Within this framework we can develop an *internal* view of spatial relations as expressed in ordinary language.

I first examine the invariance of the preposition *in*. In this and subsequent analysis the relation expressed by *in* shall be restricted to pairs of rigid bodies. Even though it is ultimately important to characterize invariance for other situations—e.g., the sugar being in the water—a good sense of the fundamental approach to invariance being advocated can be conveyed with consideration only of familiar situations with rigid bodies.

Let a and b be rigid bodies with a in b. As in Table 1 there are two cases of body b to consider. The first is when b has a closed hollow interior and a is in this interior. For simplicity of notation and analysis, I introduce the restrictive assumption that a is a spherical ball, so that the orientation of a can be ignored. Define $I(a, b) =$ the set of points in the closed hollow interior of b that the center of a can occupy. Let $\Phi(a, b)$ be the set of all transformations of $I(a, b)$ onto itself that represent possible changes in the position of a relative to b. For example, if the position of the center of a is on one side of the hollow interior of b it could be transformed, i.e., moved, to the other side of the interior without affecting the truth of the assertion that a is in b.

The set $\Phi(a, b)$ is the symmetry group (under the operation of function composition) of the position of a inside b, but because the hollow interior of b can be quite irregular in shape $\Phi(a, b)$ may not have standard properties of Euclidean isometries. It does have the expected invariance properties, namely, if $\varphi \in \Phi(a, b)$, then a with center c at point p is in b if and only if a with center c at point $\varphi(p)$ is in b.

Body b is itself subject to rigid motion in the fixed physical space. If b is closed, rotations around a horizontal axis are permitted, but if b is open, possible rigid motions of b must be restricted to those that preserve vertical orientation. And, of course, when b is subject to a rigid motion ψ, body a must be subject to the same transformation ψ, in order to preserve the invariance of the relation of being inside. Obviously, a may also be transformed by $\varphi \in \Phi(a, b)$, so that its full transformation could be the composition $\varphi \circ \psi$ with invariance of *in* preserved.

I turn now to a more extended analysis of the preposition *near*. Unfortunately, the only part of geometry that seems able to deal directly with natural objects taken as primitives is topology. Threads, knots, mazes and holes, for example, can be analyzed from a topological standpoint with considerable thoroughness, but as illustrated in Table 1, topology will not suffice even for the external analysis of many of the most common spatial prepositions such as *near*. As the sample sentences in Table 1 make plain, what at first seems to be most needed is the development of geometry based on the primitive concept of an approximately rigid body, but, as we shall see, difficulties lie in wait for this approach as well. A very preliminary and far from adequate analysis is given in Suppes (1972). It is easy to add to this earlier analysis a qualitative relation of distance between rigid bodies. The relation $ab \succeq cd$ is interpreted as the (qualitative) distance between bodies a and b being at least as great as the distance between bodies c and d. This distance is more specifically interpreted as being the minimum distance between two bodies, not the distance between their geometrical centers, which would be an alternative.

It is also feasible to introduce the affine relation of betweenness, $B(a, b, c)$. Here our intuitive qualitative judgment of body b being between bodies a and c is that something like the affine-point relation of betweenness holds for the centers of the bodies. Of course, the technical notion of the geometrical or gravitational center is not formally introduced. It is rather that we do seem to have an intuitive notion of the center of a body. Finally, the notion of one body being *near* another is relativized to context by introducing a pair of bodies a_0 and b_0 that serve as a standard in a given context. To say that two automobiles are near each other is in metric terms very different from saying that two books are near each other on the table. It is obvious that the standard for nearness changes from one context to another. Introducing the pair a_0 and b_0 is a formal device for dealing with this contextual change.

A couple of formal definitions are useful. First *a touches b* iff $ab \preceq aa$ and $a \neq b$. (I use both $\succeq$ and $\preceq$ in the standard way, as well as $\succ$ and $\prec$.) Second, *a is near b* iff $ab \preceq a_0b_0$ and a does not touch b.

DEFINITION. *A weak geometrical structure of rigid bodies* $\mathfrak{A} = (A, B, \succeq, a_0, b_0)$ *satisfies the following axioms for all* a, b, c, d, e *and* f *in* A:

A1. *The set* A *is nonempty and finite*;
A2. *If* $B(a,b,a)$ *then* $a = b$;
A3. *If* $B(a,b,c)$ *then* $B(c,b,a)$;
A4. *If* $B(a,b,c)$ *and* $B(b,d,c)$ *then* $B(a,b,d)$;
A5. *If* $ab \succeq cd$ *and* $cd \succeq ef$ *then* $ab \succeq ef$;
A6. $ab \succeq cd$ *or* $cd \succeq ab$;
A7. $ab \succeq ba$;
A8. $a_o b_o \succ aa$;
A9. *If* $B(a,b,c)$ *then* $ac \succeq ab$.

The intuitive meaning of these weak axioms is, I think, obvious. That the present setup will not lead to a metric representation without substantial modification is apparent by the failure of the qualitative additivity axiom *If* $B(a,b,c)$, $B(a',b',c')$, $ab \succeq a'b'$ *and* $bc \succeq b'c'$ *then* $ac \succeq a'c'$.

A counterexample to this axiom is easily constructed by the case of body b' being much longer along the segment joining a' and c' than is b along the segment joining a and c. The most direct way out of this problem seems to be to introduce the qualitative length of a body along the line joining two other bodies. The notation $b(ac)$ could convey this idea, i.e., $b(ac)$ is the qualitative length of body b along the segment joining a and c. We would then add such axioms as *If* $b(ac) \succeq de$ *then* $a \neq c$, and $b(ac) \succ aa$. But there are other problems to be dealt with, so I have not attempted to go further in this lecture. For example, further counterexamples can easily be constructed to the modified additivity axiom given just above even with $b(ac) \succeq b'(a'c')$ added to the hypothesis. The simplest counterexample arises from consideration of concave bodies, but without further conditions convex bodies also can be used.

But the difficulties with weak geometrical structures of rigid bodies as axiomatized is more fundamental. They do not provide the proper setting for analyzing the spatial relation of nearness or other spatial relations. The automorphisms of such a structure do not provide anything like the intuitively correct answer. What is missing is a way of expressing the many potential spatial positions of two bodies a and b, all of which potential positions satisfy the relation of nearness. We need an analysis similar to that given for *in*. Let $\Phi(a, b; a_0, b_0)$ be the set of rigid motion transformations of a relative to b such that if $\varphi \in \Phi(a, b; a_0, b_0)$ then with respect to the nearness standard a_0, b_0, $\varphi(a)$ is near to b if and only if a is near to b. As before Φ is the symmetry group, but without additional geometrical assumptions, not much can be said about its structure. Moreover, without any restrictions a and b can be subject to an arbitrary rigid motion in physical space. But in many cases it is natural to restrict a and b to a given horizontal surface, such as a table top or the floor of a room. Then the set Φ is simplified by restricting the potential changes of relative position of a to that surface. This leads on to a familiar kind of result: by considering not just one but several nearness relations of a to a number of objects we restrict dramatically the symmetry group Φ, even in extreme cases, to the identity group. But it is also clear that such exact extreme results are not part of our natural concept of nearness.

Similar analyses can be given of the internal invariance properties of the other prepositions listed in Table 1. They all have different symmetry groups, but the exact nature of each group is determined by the particular shape and size of the relevant bodies, which may vary drastically from one context to another. It is not surprising that no computationally simple theory of invariance works for ordinary spatial relations as expressed in natural language, for the robust range of applicability of such relations is far too complex to have a simple uniform geometry.

Further applications. In addition to providing an analysis of spatial terms, the kind of approach outlined above can be used in a comparison of different languages. A natural question is whether there is a universal semantics of spatial perception or do different languages have intrinsically different semantics. Certainly there are subtle differences in the range of prepositions. Bowerman (1989) points out, for example, that whereas we can say in English both *The cup is on the table* and *The fly is on the window*, in German we use the preposition *auf* for being on the table, but *an* for being on the vertical window.

Such differences are to be expected, although it is a matter of considerable interest to study how the prepositions of different languages overlap in their semantic coverage. It is a well-known fact that learning the correct use of prepositions is one of the most difficult aspects of learning a second language within the family of Indo-European languages.

One hope might be that there is a kind of universal spatial perception, so we might search for geometrical invariance across languages. But if we fit the geometries closely to individual languages, the primitives but not the theorems may be different in the sense of having a different geometrical meaning. A related set of questions can be asked about the order of developmental use of spatial prepositions by children. This seems to be a particularly good case for careful examination of what happens in different languages, because of the relatively concrete and definite rules of usage that govern spatial prepositions in each language in their literal use. Bowerman (1989) has an excellent discussion of many details, but does not focus on systematic geometrical aspects.

2. GEOMETRY OF VISUAL SPACE

Let me begin by reminding you of the classical alley experiments of Hillenbrand (1902) and Blumenfeld (1913). Almost all students of psychology at least know something about these classical experiments. I shall mainly refer to Blumenfeld's work because it was an improvement on that of Hillenbrand. As you will perhaps recall, Blumenfeld performed experiments with so-called parallel and equidistance alleys. The subject sits at a table in a darkened room. Looking straight ahead, he is asked to adjust two rows of points sources of light placed on either side of the normal plane, i.e., the vertical plane that bisects the horizontal segment joining the centers of the two eyes. The two furthest lights are fixed and are placed symmetrically and equidistant from the normal plane. In the case of the task being the construction of a parallel alley, the subject is asked to arrange the other lights so that they form two parallel lines extending toward him from the fixed lights. The subject's task is to arrange the lights so that they are perceived as lying on parallel lines in the subject's visual space. The other task is to construct an equidistance alley. In this experiment, all the lights except the two fixed lights are turned off and in sequence a pair of lights is presented, which are adjusted to be at the same perceptual distance apart as the fixed lights. Here the subject is making a clear judgment of equidistance, not of lines being parallel. When one pair of lights is turned off, another pair closer to the subject is presented for adjustment, etc. The important point now is that the physical configurations arising from the two experiments do not coincide, but in Euclidean geometry straight lines are parallel if and only if they are equidistant from each other along any mutual perpendiculars. Classically, the discrepancies observed in the Blumenfeld experiment are taken to be evidence that visual space is not Euclidean. The results are shown graphically and thereby most easily in Figure 1. In both the parallel alley and equidistance alley experiments, the lines are found to diverge as the adjusted pairs of points lie further away from the subject. But the angle of divergence tends to be greater in the case of parallel than in the case of equidistance alleys, as is clear in Figure 1. Since the most distant pair of points is the same for both alleys, this means that the equidistance alley lies outside the parallel alley. These results have been taken by Luneburg and others to support the hypothesis that visual space is hyperbolic, for this qualitative result is a property of hyperbolic space, even though there is some ambiguity in the fact that to a given line there is not a unique parallel line in hyperbolic space. Luneburg essentially used orthogonality to characterize being parallel, a matter that is discussed with some care in Indow (1979).

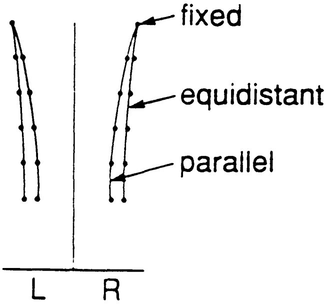

Figure 1. Diagram of classical alley experiments.

Although the experiments and Luneburg's conclusions are well known, the situation is not conceptually as clean as it might be for, as you will remember, the alternative to the space being Euclidean or hyperbolic is that it is elliptic—the restriction to these three choices will be discussed more in a moment. However, no two lines can be parallel in elliptic space, so again a compromise must be struck in how the notion of parallel is to be handled. However, it must be said that for the local concept of two lines being parallel, it can be shown that in elliptic spaces the parallel alley lies outside the equidistance alley.

If matters were simply to be left with the classical alley experiments, which have been duplicated many times and have practically become a standard demonstration experiment, then we could settle the issue quickly, by concluding the evidence was excellent that if we must choose between the Riemannian surfaces of constant curvature in characterizing visual space, hyperbolic space is obviously the correct choice. However, as in most such matters in perception, the subsequent history of new and different experiments has ruled out any simple conclusion.

Before looking more systematically at some parametric experiments, it is important to note that the experiment that meets the criticisms given above of the notion of parallel is that of Blank (1961) who asked subjects to compare line segment *bc* and line segment *ef* in Figure 2. I am not describing the exact experimental protocol, but the point was to get judgments from subjects as to whether *ef* was half of *bc* (Euclidean space),

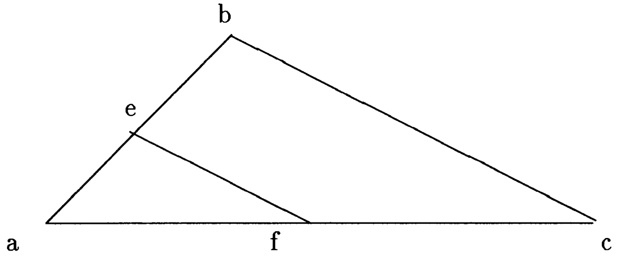

Figure 2. Illustration of comparisons required in Blank experiment.

less than half of *bc* (hyperbolic space), or more than half of *bc* (elliptic space). A majority, but not all, of the subjects supported the hyperbolic hypothesis. From a methodological standpoint this is in my judgment a very nice experiment, even though there are some natural qualms about judgments of one segment being twice another, as compared with a more direct qualitative judgment of equidistance or parallelness, or, to put the matter another way, in some fundamental geometrical sense the notion of parallel or equidistance is more fundamental than that of being half the length. This is a minor objection however, and I think it is overridden by the elegant way in which the problems of parallelness for hyperbolic and elliptic spaces are avoided.

Luneburg has been the central theorist of the hyperbolic conception of visual space and he has well worked out theoretical ideas that have led to a number of experiments (see for example his publications of 1947, 1948, and 1950). His central idea was to develop a parametric theory based on the general assumption that in order to have free mobility of rigid bodies the space must be a Riemannian space of constant curvature. Luneburg used a somewhat unsatisfactory differential argument to get the space of constant curvature and did not refer in a detailed way to the classical Helmholtz-Lie space problem which concerns the characterization of spaces that satisfy free mobility. I shall not review that history here but only recall for purposes of present reference that the results came out as Luneburg had hoped. The only spaces tolerating free mobility of rigid bodies are Riemannian spaces of constant curvature if we demand satisfaction of certain other natural properties such as smoothness.

Luneburg showed that the line element ds can be expressed in terms of orthogonal sensory coordinates ξ, η and ζ by:

$$(1) \qquad ds^2 = \frac{d\xi^2 + d\eta^2 + d\zeta^2}{[1 + \frac{1}{4}K(\xi^2 + \eta^2 + \zeta^2)]^2}$$

where for:

$$\begin{array}{ll} \text{Euclidean space}: & K = 0, \\ \text{hyperbolic space}: & K < 0, \\ \text{elliptic space}: & K > 0. \end{array}$$

By introducing certain relatively natural but restrictive psychophysical assumptions, Luneburg shows that in physical coordinates Equation (1) becomes in polar coordinates

$$(2) \qquad ds^2 = \frac{4}{(e^{\sigma\gamma} + Ke^{-\sigma\gamma})^2}(\sigma^2 d\gamma^2 + d\varphi^2 + \cos^2\varphi d\theta^2) \quad ,$$

where on the basis of the psychophysical assumptions the relation to the sensory coordinates is postulated to be the following:

$$\begin{aligned} \xi &= 2e^{-\sigma}\cos\varphi\cos\theta \\ \eta &= 2e^{\sigma}\sin\varphi \\ \zeta &= 2e^{-\sigma}\cos\varphi\sin\theta. \end{aligned}$$

The parameter γ has a physical definition, but the parameters σ and K are estimated for each subject individually. The individual subject estimates for these two parameters, especially for K, is a reflection of the fact that the specific curvature observed by any two subjects, even if both observe hyperbolic space, will in general be different.

Many experiments have been done within this Luneburg framework. An early study is Hardy, et.al. (1953), but by far the most sustained experimental program has been that of Tarow Indow and his collaborators beginning in 1962 (Indow 1967, 1968, 1974a, 1975, 1979, 1982; Indow, Inoue and Matsushima, 1962a, 1962b, 1963; Matsushima and Noguchi, 1967; and Nishikawa, 1967).

I shall not attempt to summarize these many carefully performed experiments in any detail. (For more detailed summary, see Suppes, et.al., 1989, pp. 145–153.) The most relevant three conclusions are these: (i) For most subjects $K < 0$, (ii) the estimates of K and σ were very unstable for many subjects even when the same experimental conditions were repeated, (iii) values of K and σ did not transfer well from one experimental setup to a different one. In particular, attempts to transfer K and σ from one set of experiments to the alley experiments did not work well at all.

The experiments I am summarizing in cursory form are in my judgment among the most careful of any psychological experiments involving parametric estimation that I can think of. The inference is rather about the unsatisfactory character of the theory rather than of the nature of the experiments. The great instability of the estimated parameters, especially for conceptual purposes the great instability of K, is in marked contrast to the precision with which the parameters of physical space are measured, with uniform values holding over a great range of circumstances. The instability and lack of generalizability naturally generate skepticism that it is the right scientific move to think of visual space in the same kinds of terms and within the kind of conceptual framework so common in examining the nature of physical space.

Moreover, there are several experiments that raise further doubts, beyond any question of instability of parameters, because they take the results for the nature of visual space outside the Luneburg framework. The first experiment to be mentioned is that of Wagner (1985). The methodological approach of this experiment is notable for two reasons. First, unlike the standard Luneburg experiments, the experiment was conducted outdoors in full daylight in a large field with subjects making judgments about the geometrical relations of 13 white stakes. Different procedures were used for measuring distances, angles, and areas. In particular magnitude estimation, category estimation, and constructing a simple scale map were used for judging distances, the only results to be considered here.

The results are extremely interesting and are contrary in important ways to essentially all of the Luneburg-type experiments. The important result is that there was spectacular foreshortening in depth perception. Let the x axis be the horizontal depth axis, that is, the axis perpendicular to the vertical plane through the eyes, and let the y axis be the horizontal frontal axis passing through the two eyes. Let two physical distances be such that $x = y$, that is, one distance is taken along the depth axis and the other along the frontal axis. Then in perceptual estimates (indicated by primes) of depth $x' = 0.5y'$ with of course some variation around 0.5 for individual subjects. The coefficient 0.5 is not some minor variation on standard physical Euclidean space but a major deviation in the form of an affine transformation of Euclidean space. It would be extremely interesting to determine if perceptual physics suffers such a large affine transformation as perceptual geometry. There have been other experiments reporting such depth foreshortening, for example, Battro, Netto, and Rozestraten (1976) but none as striking as the experiment I now turn to.

This is an elegant older experiment by Foley (1972) which leads to even stronger results, results that require the conception of visual space to lie outside any of the Riemannian spaces of constant curvature. In fact outside of any of the geometries ordinarily used in the study of perception. In Foley's 1972 experiment, the subject sat in a dark room with both eyes open, and a light a was fixed in the subject's visual field in the horizontal

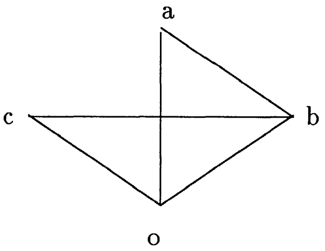

Figure 3. Illustration of comparisons required in Foley Experiment.

plane at the eye level. The subject was asked to set light *b* so that the line *ob* (*o* is the position of the subject) was perceptually perpendicular to line *ab*, and segment *ob* was apparently equal in length to segment *ab* (see Figure 3). Notice that this kind of task is very similar to the sort of task arising in the classical alley experiments, as far as the judgments required from the subject are concerned. The subject was next asked to set light *c* so that *oc* was perceived to be perpendicular to *ob* with *oc* equal in length to *ob*. The subject was then, as a final task, asked to judge the relative lengths of *oa* and *bc*. The important point is that for homogeneous Riemannian space, whether it be Euclidean, hyperbolic, or elliptic, by construction the right-angled isosceles triangles *oba* and *boc* should be congruent, and so *oa* and *bc* should be judged equal in length. The experimental results were quite different, however. Twenty-four subjects in 48 trials judged *bc* to be significantly longer than *oa*. It is important to note about Foley's experiment that this is not a question of various symmetrical, contextual features being present that lead to distortion. The general experimental environment is essentially that of the standard Luneburg experiments, or the standard alley experiments. The results however, are disastrous for any simple geometrical theory of visual space, for it requires us to move outside the framework of the standard elementary homogeneous spaces. I want to turn now to what these various results suggest for our study of visual space. I have organized the remarks under several different headings.

Study of qualitative axioms. Foley's experiment as well as other experiments he has conducted on the verification of Desargues' theorem, point toward a more intensive program of experimentally testing which individual qualitative axioms are satisfied. I say "which" with the understanding that such investigations could well begin with the standard classical primitives of geometry. For example, the linear relation of betweenness, standard qualitative relation of congruence, standard perpendicularity and parallel relations, etc. On the other hand, given the complicated and subtle nature of the results it may be that different primitive concepts will turn out to have better invariant properties. For example, if it is hard to get coherent congruence results as we rotate 90 degrees from the frontal to the depth axis, it is natural to think that a different and more complex notion of congruence is needed than the standard one, which is independent of orientation—I return to this idea in a moment.

Space has not permitted me to say anything about projective geometry here, but the recent book of Cutting (1986) provides an excellent overview of projective questions that are natural to ask about visual space. The point of mentioning projective geometry here, however, is to remark that it may be that the kind of finite geometry characteristic of earlier work in projective geometry and still used for counterexamples and for other purposes, may turn out to be something that needs to be studied in the present context, for it may be that we would be able to satisfy certain qualitative axioms for a finite set of points as in the Foley experiments, but not their extension to a larger number. There is good reason why experimenters have been reluctant to move in this direction, for the possibilities of finding finite spaces that will satisfy a given fixed set of points are many, and it would be easy to get some extraordinarily ad hoc results which would not be of much interest. It seems to me, however, there are certain principled lines of inquiry that could be used and that might produce some very interesting conceptual results about the visual space of these particular experimental configurations that produce results that are so difficult to interpret. I also want to emphasize another way of thinking about these finite spaces that is different from the way finite projective spaces were thought about in the past. It seems to me the right way to think about them is in terms of geometrical *constructions*. In this case we can of course think of satisfying a much larger set of points by conceiving of the experimental configurations as being the first steps in constructing an ever larger configuration of points.

In the Foley experiments, for example, the subject has two fixed points (see Figure 3), namely points a and o. The other points are generated by construction. For the constructions involved, the line segments cut the depth axis and frontal axis all at a 45^o angle, and this could be a qualitative restriction on constructions leading to congruence. The observant reader will have noted that Foley's experiments satisfy the foreshortening result of Wagner's experiments. What Foley's results in conjunction with Wagner's show is that we cannot have some simple concept of congruence resulting from an affine transformation of the Euclidean space. We need something like a "directional" concept of congruence.

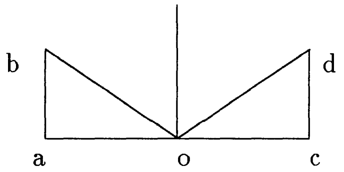

Figure 4. Illustration of comparisons permitted by restricted axiom of congruence

Construction of a small number of points, with direction a part of the primitive concept, and also a part of the primitive concept of congruence, could lead to qualitative axioms producing both Foley's and Wagner's results.

Here is a sketch of one approach that might work. First, we use some standard qualitative primitives for affine spaces, e.g., betweenness or parallelness. Second, we develop in this framework affine congruence, i.e. congruence restricted to parallel segments. Third, add a restricted congruence axiom of the following sort that takes account of direction by requiring symmetry about the depth axis. (Here, as before o is the position of the observer.)

AXIOM. *If $B(a, o, c)$, $ao \simeq oc$, ab and cd are parallel to the depth axis, and $ab \simeq cd$, then $ob \simeq od$.*

Note that the congruences postulated in the hypothesis of the axiom are affine, but the congruence of the conclusion is not. Figure 4 shows the simple construction. Other standard axioms of congruence that would be assumed are the following:

1. *If $aa \simeq bc$ then $b = c$.*
2. *$ab \simeq ba$.*
3. *If $ab \simeq cd$ and $ab \simeq ef$, then $cd \simeq ef$.*
4. *If b is between a and c, b' is between a' and c', $ab \simeq a'b'$ and $bc \simeq b'c'$, then $ac \simeq a'c'$.*

The axioms of congruence given are not enough to prove a representation theorem for the kind of space suggested by the Foley and Wagner experiments, but they do provide a variety of testable consequences about congruence that are not falsified by the Foley and Wagner results.

Context effects. Unfortunately, too energetic an effort to give a very detailed qualitative theory of the Foley and Wagner type of experiments could be misplaced, because already different results in a not too dissimilar arrangement are obtained in the classical alley experiments, which, at least in the obvious interpretation of the constructions made by the subjects in the experiments, do not satisfy the affine properties postulated for the Foley and Wagner experiments. This is perhaps one of the most disturbing aspects of experiments on visual space, namely, different experimental configurations can produce different geometrical results. I have made the point on several occasions that it may be the case that classical geometry is the wrong model for visual space. The kind of contextual effects to be seen in different arrangements is something much more characteristic of physics than geometry. In classical geometry there are no context effects. The properties of a configuration are not affected by properties of neighboring configurations. But in physics it is quite the opposite. When we study the interaction of two bodies we expect something very different to happen if we change the context by introducing a third body. Essentially every significant theory in physics has this kind of contextual property. What is disturbing, however, is the apparent difficulty of analyzing context in visual experiments in a way that would lead to interesting systematic results. To put the matter another way, in the framework of a central thrust of this lecture, we have not yet been successful in finding general principles of visual perception that have the appropriate invariance properties. It is in fact an open question whether satisfactory general principles exist. Our perceptual apparatus may be a pluralistic assembly of systems without strong unifying principles.

31

CAN PSYCHOLOGICAL SOFTWARE BE REDUCED TO PHYSIOLOGICAL HARDWARE?

The question of the title I answer in the negative. There are four strands to my argument. The first, which corresponds to the first section of the paper, analyzes the nature of computation. The second concerns the nature of goal-oriented behavior. The third uses an argument that the mind is computationally irreducible. The fourth asserts the irrelevancy of the standard attempts to provide a reduction via general ideas about determinism.

1. NATURE OF COMPUTATION

We may stipulate, I believe, for this paper that the mind is among other things a computational device. This means that matters of computation are of central importance in any arguments about the reduction of psychological concepts to physiological ones. Part of my argument about the irreducibility of computational concepts of the mind to physiological concepts is from the much simpler case of digital computers. In the case of digital computers, we understand to a very much more thorough

*Reprinted from *The problem of reductionism in science* (ed. by E. Agazzi), 1991, pp. 183–198. The Netherlands, Dordrecht: Kluwer.

degree exactly the physical basis of computation and at what level the computational concepts interface the physical concepts. In spite of this great precision of knowledge of interface, we do not at all attempt in the standard theory of computation for digital computers to replace computational concepts by physical ones, which corresponds to replacing psychological concepts by physiological ones. In fact, I am skeptical that even in the case of relatively simple digital computers we could make *direct* physical observations on the computer from which we could infer in detail and without any high degree of error what from a software standpoint was actually being computed. The situation is very much more complicated and difficult, and less likely ever to be understood thoroughly, in the case of our own mental computations, because there is no evidence we will make much headway on the detailed physiological or physical identification of the neurons that are doing any particular computation. Notice of course that this lack of clear identification of physical location, quite apart from understanding the details of that physical location is characteristic of computation in a digital computer. It is a day dream to think that we can easily identify where a particular computation is taking place. Computations in a modern computer move around dynamically. Where they are even placed initially is not a static concept but a dynamic one depending upon what is present and what else is being computed at the time computation is started. It would be a great surprise if something similar is not true of the computations in the brain. Location of mental computations in terms of individual neurons seems totally out of reach. Global location of computations of a particular kind being done in a particular region of the brain is sometimes feasible.

It is also important that physically very different computers compute the same function even by different software programs. Only isomorphism at a high level is usually of interest, and really never in terms of concepts of computation at the level of individual transistors, or to be even more reductionist, at the level of individual elementary particles which make up the many different microscopic parts of a transistor.

Still another concept that seems likely to hold for the brain is that neurons compute statistically, unlike most of the current digital computers. With a statistical computation it is especially unfeasible to think about a reduction from the software to the hardware. Each neuron is making a statistical contribution, but the physical performance of a particular neuron is not of any decisive importance. The feeling is rather here that we have something like the standard result for random variables in probability theory. Given the random variables, which roughly speaking are meant to correspond here to the software concepts, we only have requirements of consistency for there to exist a common sample space.

The sample space is never unique. The basic consistency theorems, for example, Kolmogorov's theorem, deal with the existence of an underlying sample space. Without a very large number of assumptions that are not part of the standard theory, there is no unique sample space. The same is surely the situation with the statistical computation of neurons. The underlying neurons actually making the computation can in all likelihood never be observed and certainly not with present concepts for observation of neurons. A given computation in the brain may be located in different neurons depending upon the dessert one had for lunch or the last green perception that passed through the system. There is such intensive contextualism that a reasonable conjecture is that identifiability for purposes of reduction is out of the question.

In order to eliminate fairy tales, it is important to tighten the argument here and to say that the formal claim of reduction not being possible is relative to a set of observational variables and observational techniques. I will not in this paper attempt such a formalization but I certainly think it is possible and can be done in a straightforward way for simple examples. Obviously, I am not intending to give an *a priori* argument that will hold for all time regardless of what scientific methods are in place a thousand years from now. I am concerned to give at a foundational level an argument in terms of current science and relevant philosophical concepts, an argument that is meant to be a strong one from a computational standpoint, against any possibility of reduction relative to any set of currently observable concepts.

There is one point that might seem elusive in making the distinction between software and hardware in the case of biological organisms as opposed to digital computers. Of course, even in the case of computers, the distinction is not as sharp as it might seem, for in some sense the software program must become a part of the hardware, i.e., a part of the physical organization of the computer. Where, it might be asked, does hardware stop and software begin. Once the software is embodied in the computer, as it is in a different way in the brain, this is not an easy question to answer from purely physical considerations in the case of the computer or physiological ones in the case of the brain. We are able to make the separation in the case of digital computers only in terms of knowledge of what has been done in a deliberate fashion to program the computer, and how the computer hardware has been organized to embody programs. Even this formulation for digital computers is much too simple. First, as already pointed out, the physical location of particular pieces of the program in the memory and processors of the computer is a matter of dynamic allocation and not something we can physically easily directly describe. More importantly, with the current emphasis on digital computers acquiring the capability of learning, the details of the program will not necessarily always be possible to identify.

At least at present our main way of thinking about the brain's software is just in terms of the kind of mental or behavioral concepts psychologists have been developing for a long time and the language of common experience for a much longer time. Moreover, it is undoubtedly these concepts that are the significant targets of any reductionist program. There is no doubt a more detailed and extended sense of software that could be defined. We could include the structural and functional computational changes to be attributed to learning rather than genetic inheritance. Whichever view is taken, reduction to the brain's hardware has no present hope of being carried out in detail.

2. GOAL-ORIENTED BEHAVIOR

A second strong argument against reduction of psychology to physiology is to be found in goal-oriented behavior in humans and other animals. If I ask my well-trained dog to fetch the newspaper, which has just been delivered, no amount of physiological or physical observation of the dog would be able to either predict the trajectory of his motion as he goes in search of the paper, or infer what task he was attempting to carry out. I am not suggesting that the activity of the mind operates without use of the brain, it is just that psychological concepts cannot be reduced to physiological ones. In other words, we need psychological concepts and the theories that embodies these concepts as theories that can be proved independent from a logical standpoint of purely physiological theories. It is a scientific fantasy to think we shall ever be able to make within our present scientific framework sufficient observations on my dog or on any other to determine what task it is engaged in, if those observations are restricted to purely physiological, including neurological, methods of observation and analysis.

To draw a drastic parallel—perhaps too drastic for some—, consider the use of logic in formulating physical theories. No one would be so foolish as to say that we can reduce physics or physical concepts to a matter of logic or logical concepts, just because we need logic in the formulation of physical theories. This is how I see the relation of psychology and physiology.

To talk about the brain as a machine that we can understand mechanically, which is familiar talk among physiologists and even neurologists, is to talk in a mistaken way. Think how absurd it would be to hear physicists talking about physics as a logical subject and meaning by that that only logic was needed to formulate physical theories. (It is as if logical methods for rigorously proving the independence of axioms or concepts had not been developed for the past hundred years.)

There is another point to be made about goal-oriented behavior. It is easy enough for almost everyone to accept the fact that from purely physical or physiological observations no one can predict where I am going as I leave the house on a complicated physical trajectory to my office, to another house, to a store, or to a restaurant. However, it could be claimed and would be by some die-hard reductionists, that this is simply a case of unpredictable behavior, also to be found in purely physical systems, but unpredictable behavior that can be explained by purely physical concepts. The separation of explanation from prediction is an important matter scientifically but doesn't really bear on the present case, in the following sense. No matter how many observations or how much information of a purely physiological or physical sort is to be collected prior to or after I execute my chosen route to restaurant, store, or whatever, it will be impossible on purely physiological or physical grounds to explain the complicated path I followed. The man in the street recognizes such an enterprise as nonsense. It is important to recognize that from the most fundamental scientific standpoint it is nonsense as well to think in terms of being able to make such a purely physiological or physical analysis of my or any other higher organism's complicated movements.

Physiologists sometimes talk about goal-oriented behavior in cells. Without attempting to judge the scientific merit of this line of analysis, it is evident that we do not have the faintest idea of how to reduce the goal-oriented behavior of a higher organism to goal-oriented behavior of its individual cells. In other words, global goal-oriented behavior cannot be successfully analyzed in terms of goal-oriented behavior of cells. There is a seductive analogue here that could lead to mistaken conclusions. The analogue is that of the reduction of thermodynamics to statistical mechanics. In this case, the behavior of macroscopic parts of matter is reduced at a physical level for certain concepts to the behavior of microscopic particles. Moreover, the relation, as suggested above for neurons in mental computation, is statistical. However, the enormous difficulty of making this reduction a rigorous one, even under the simplest conditions, shows how improbable it is that at any time in the foreseeable future of science we would have the faintest idea of how to carry through a serious reduction of global goal-oriented behavior of an organism to behavior of the organism's individual cells. There is something enormously seductive about this analogue. It is natural to think that we should somehow be able to push through a program of reducing our ordinary behavior as persons to the structure and function of the many billions of cells that make up our bodies. It is a metaphysical dogma that is hard to dislodge. My point is not to say that I can prove it as false, but just to say that the evidence for it is negligible. A way of putting the point is that it is very unlikely that the concepts needed to describe the global behavior of an organism can be reduced to concepts applicable only to individual cells.

It is important to block one mistaken conclusion from the argument I have just given. What I have said is not meant to suggest that we cannot identify at a level even below that of individual cells, for example, at the level of DNA or at the level of genes, microscopic features that predict major features of behavior. The triumph of genetic analysis of many different sorts of diseases is a major triumph, and something very special about science in this century. It is, on the other hand, idle to think that we have even begun to touch the problem of being able to infer goal-oriented movement of an organism from cellular observations. The usual scientific pluralism is at work here. We can do some things well, but not others, in terms of reducing global aspects of behavior to molecular ones. It is a form of metaphysical imperialism—in my view a scientific mistake—to think that we can generalize the successes of molecular biology to carrying out anything like a reduction of all major aspects of goal-oriented behavior. I mention again the difficulties that have been encountered in the last two decades in carrying through in a rigorous way the program of reduction of thermodynamical systems to statistical mechanical ones. It is easy to give from the current literature examples of thermodynamical systems that we do not know how to reduce to statistical mechanical systems. The incomparably more subtle and difficult problem of reduction of the theory of movements of higher organisms seems scientifically totally out of reach.

3. COMPUTATIONAL IRREDUCIBILITY

A familiar and important, but not always remarked upon, property of classical physical systems that have been the object of much attention in the history of physics is the property of being computationally reducible. Here is the simplest and most important example. Newton's solution of the two-body problem, i.e., the problem of motion of two bodies acted upon only by the forces of their mutual gravitational attraction, permits us to predict the motion in the future or the past, or the position at any future or past time, given appropriate data on initial conditions at a given time. Moreover, this model has the important property of being applicable in first approximation to two-body systems for the planets, with the sun as one of the bodies. The fact that we can solve the equation of motion in closed form and thereby compute quite directly the future path of the bodies is of fundamental importance. Unfortunately, there was for a long time the feeling that this would be the norm for physical systems, i.e., that most of them were computationally reducible. We would be able to solve the equations of motions to determine the paths of the particles for any indefinite time into the future. However, already in the nineteenth century, the intractable problem of moving from two bodies to three bodies gave plenty of evidence that our ability to computationally reduce most physical systems was probably extraordinarily limited. Moreover, even when we cannot solve the equations of motion in closed form, we often feel that we can do a very good job of a numerical approximation. Already, however, in the case of the three-body problem, as was essentially shown by Poincaré, this was not possible. Now we understand the phenomena very thoroughly for systems that are drastically unstable, as is the situation for some initial conditions in the three-body problem. The numerical methods of computation, necessarily approximate in character, provide a very limited horizon of practical computation concerning the behavior of the system or, to put it in other terms, a very limited horizon of predictability for the behavior of the system. Many other systems of a similar simple physical character have now been identified. With the modern intense interest in chaotic systems we have a sense of limitations in principle about predictability and about computational reducibility of physical systems that did not exist until rather recently, even in so well-established an arena as that of classical mechanics.

There are many reasons to think the mental computations of the mind are also computationally irreducible. One consequence of this is that we shall not be successful in simulating artificially the behavior of the brain. We shall not be successful in the sense that important aspects of human behavior will be missed in any such simulation. Even a model of ten billion artificial neurons will be deficient in providing anything like predictive or computational models. Notice that what we would like is really hopeless: speed up the computations of the brain by four or five orders of magnitude with a model of three or four orders of magnitude less neurons and thereby predict rather well by such computational reducibility future behavior. An unlikely story if ever there was one. To argue that the mind is computationally reducible, as in the case of arguments about physical systems, does not mean that we cannot find subsystems or aspects of behavior that can be computationally reduced. In other words, we can make certain theoretical computations about the behavior of the system that can be verified, and the computation is much simpler and faster than the behavior of the system itself. For example in the case of the three-body problem we can compute for restricted cases the escape velocity of one body, and we can make predictions about the behavior of a qualitative sort without being able to make computations about the detailed behavior. But just as in the case of such physical systems, the brain cannot be computationally simulated in simpler fashion, i.e., be reduced to the computations of the simpler system, when we are concerned with its full behavior. It is a piece of unrealizable scientific fantasy to think that we can move our minds to simulated brains and preserve our psychological identities. There is no reason whatsoever to think such a computational transfer will ever be possible. Above all, the physiologists and neurologists will never make a computational reduction to formulas that lead from individual cell behavior to the mental computations of a fully functioning brain. I am not proposing to offer a metaphysical proof that such a neurological reduction is impossible, it is just that there is no serious scientific evidence whatsoever that it ever will be achievable within the framework of science as we now conceive it.

If I am right in this last claim, it means that in any serious formal or scientific sense reduction of psychological concepts to physiological ones will not be possible. Here what I state informally I mean in a more formal way. Given a formally and empirically adequate psychological theory of psychological phenomena, it will not be possible to prove a representation theorem in terms of a formally and empirically adequate physiological or neurological theory. A certain kind of handwaving may be indulged in by reductionist-minded philosophers, but no serious demonstration of reduction will be given, and it will not be given for substantial reasons. There is no scientific evidence that such a reduction can be carried out at a satisfactory level of detail.

There is still another and different point to be made about computational irreducibility. If we think of having a uniform theory of neurons—meaning that neurons act in the same way from one individual to another and their interaction with software is the same, reduction seems unfeasible. The hopelessness of the situation increases even more when we introduce the hypothesis, which seems likely to have considerable support, that the way in which individual neurons in a given individual interact with software is different from person to person. This would be a natural consequence of biological development occurring in a partially random fashion at the level of dendritic formation and the learning experience in terms of which some of that development is influenced also occurring with random variation from one individual to another. This would mean that the hardware of the neurons is connected to the software of thought in quite different ways in different individuals. If this conjectured variation from individual to individual holds, then reduction is all the more impossible. The detailed structure of computations in one individual, taking the hardware and software together, would differ in quite significant ways from the corresponding structure in any other individual.

4. IRRELEVANCE OF PHYSICAL DETERMINISM

The argument for reduction of psychology to physiology, as a byproduct of the reduction of physiology to physics, as a consequence of determinism, is usually not put in as direct and simple a way as I will put it here. I think, however, the force of the argument is as follows. The physical universe is deterministic because as some analytic philosophers, untrained in physics, would put it, it is analytic that like events must have like causes. Given that there is a deterministic account of the physical universe, it then follows that everything that takes place in that physical universe is equally determined. If we know all there is to know about the physical world, that will fix uniquely all the other phenomena including, of course, the mental activities of higher organisms.

Philosophers have been ringing the changes on this argument with different degrees of explicitness and different degrees of emphasis for a long time, at least since the appearance of Kant's *Critique of Pure Reason* in the latter part of the eighteenth century. Kant doesn't discuss explicitly the concept of determinism, for it had not really surfaced in a completely clear way, even though there is a famous passage about the deterministic nature of the universe in Laplace's introduction to his treatise on probability which appeared not long after the publication of Kant's *Critique*. But there is no doubt that Kant implicitly adopted a deterministic view in his using classical physics in generalized form as the metaphysical foundation of natural science and in his treatment of the category of causality in the *Critique of Pure Reason*.

Of course, Kant was not for a moment prepared to adopt the view that psychology could be reduced simpliciter to physical determinism. As in his treatment of the antinomy of free will, he was quite prepared to bite the bullet and remove psychology entirely from the domain of science, or as he would put it, more restrictively, natural science. The radical character of Kant's solution to the acceptance of physical determinism, as I would put it, has not always been properly recognized when it comes to working out what one could then hope to do with the science of psychology, but he certainly lays out his views in as explicit a way as could be asked for in the following passage in the preface to the *Metaphysical Foundations of Natural Science*:

> But the empirical doctrine of the soul must always remain yet even further removed than chemistry from the rank of what

> may be called a natural science proper. This is because mathematics is inapplicable to the phenomena of the internal sense and their laws, unless one might want to take into consideration merely the law of continuity in the flow of this sense's internal changes. But the extension of cognition so attained would bear much the same relation to the extension of cognition which mathematics provides for the doctrine of body, as the doctrine of the properties of the straight line bears to the whole of geometry. The reason for the limitation on this extension of cognition lies in the fact that the pure internal intuition in which the soul's phenomena are to be constructed is time, which has only one dimension. But not even as a systematic art of analysis or as an experimental doctrine can the empirical doctrine of the soul ever approach chemistry, because in it the manifold of internal observation is separated only by mere thought, but cannot be kept separate and be connected again at will; still less does another thinking subject submit to our investigations in such a way as to be conformable to our purposes, and even the observation itself alters and distorts the state of the object observed. It can, therefore, never become anything more than a historical (and as such, as much as possible) systematic natural doctrine of the internal sense, i.e., a natural description of the soul, but not a science of the soul, nor even a psychological experimental doctrine. This is the reason why in the title of this work, which, properly speaking, contains the principles of the doctrine of body, we have employed, in accordance with the usual practice, the general name of natural science; for this designation in the strict sense belongs to the doctrine of body alone and hence causes no ambiguity (Kant, 1970, pp. 8–9).

Unfortunately, not many philosophers, and I would say almost no scientific psychologists, would be prepared today to follow Kant's way out of the dilemma of determinism.

Of course one immediate response to what I have called the dilemma of determinism is the modern one of saying that quantum mechanics has shown that the microscopic world of physics is not deterministic. I am not going to take that line of argument here because I don't believe it. My own view about quantum mechanics, expressed in several places, is that quantum mechanics is a weak probabilistic theory of the mean (Suppes, 1990). In this view, quantum mechanics is compatible with both deterministic and indeterministic hidden-variable theories, of which perhaps the best example of the latter is stochastic mechanics as developed by Edward Nelson and others, with the understanding that classical Markovian assumptions of Brownian motion must be relaxed to deal with problems of locality. But also as a viable possibility is the kind of deterministic hidden-variable theory outlined by David Bohm and various colleagues. I do not claim fully to understand Bohm's ideas and they are yet to be given an articulated and detailed development, but there is no reason to think that they cannot in principle be elaborated. The difficulty, of course, with any of these extensions of quantum mechanics, in the sense of providing a hidden-variable theory,—one that takes proper account of locality problems—, is being able to make an experimental determination as to whether or not the theory is correct. Such theories may get defeated in the second round by their attempts to go beyond the phenomena of classical quantum mechanics to relativistic particle phenomena, quantum electrodynamics, and more generally to the wide range of experimental findings in elementary particle physics.

What I want to argue is that whoever is right about the proper hidden-variable theory for quantum mechanics, which may turn out to be a purely metaphysical choice, determinism as a general thesis is irrelevant to the question of the reducibility of psychology to physiology. The reason for my holding this view is easy to state. Determinism is too capacious and general a theory to help any such issue to be settled in an interesting way. Why is this? Because the collection of theories that are deterministic is able to accommodate any sort of behavior. Perhaps the way to illustrate this without too many complications and reservations is to consider again the physically simple case of the three-body problem that I have discussed elsewhere with reference to propensity theories of probability (Suppes, 1987). As we move from two-bodies to three-bodies, our detailed understanding of the motions of the three bodies disappears, a fact, which I pointed out earlier, has been well-known since the nineteenth century. What has not been well-known since the nineteenth century is the proof that for rather simple restricted cases of the three-body problem—meaning a reduction of the problem to the motion of a single-body where the motion of that body is determined by the other two bodies—the following sorts of results hold. First, there exist initial conditions, which in this case are just the initial position in one-dimension and the velocity in one-dimension of the body, such that the sequence of the largest integer values contained in the temporal half-cycles of passing through the plane of the other two bodies has the following property. The sequence of integers so generated, the so-called symbolic dynamics, can represent any random sequence of integers, where the integers are greater than a certain constant. We can therefore represent in terms of the symbolic dynamics of this simple deterministic system—simple in terms of understanding its causes and the derivation of the differential equation governing its motion, not simple in terms of its actual motion—, any random sequence of heads and tails. Second, in contrast there exist initial conditions, in this case just a single number, the velocity in one-dimension of the body, such that the symbolic dynamics encodes the contents of the books in the Library of Congress. One example is purely random, the other is as intentional as you wish, but the richness of this simple, deterministic system is capable of generating either phenomena.

As I have argued in another paper we can develop the same line of attack with purely indeterministic systems (Suppes, 1991a). If you don't like determinism choose indeterminism. This choice corresponds to the two choices of hidden-variable theories for quantum mechanics I mentioned above.

So does one choose between indeterministic structures of some general probabilistic theory or unstable structures of some deterministic theory? Put in very broad terms the choice seems to be a matter of taste in metaphysics. At the present stage of science there seems no likelihood of any sequence of crucial experiments that will force one of the two positions to the wall. Indeterminism and determinism are here to stay. Exercise your metaphysical choice as you will. There is no inconsistency between determinism and randomness. We can use unstable deterministic systems to generate any probabilistic phenomena desired, or we can take a system that is indeterministic and not known to be deterministic, if that is your metaphysical bent. There are, in fact, some beautiful theorems by Donald Ornstein and his colleagues that make the metaphysical point in still stronger fashion: there are physical systems on which we can make an infinite number of observations—or if you want more precision, there are mathematical models of certain physical phenomena—, such that on the basis of an infinite number of observations it is impossible to distinguish between a deterministic mechanical model governing the phenomena and a stochastic process governing them. This line of argument, the last line of argument I consider here, can be summed up in this way. The classical attempt to reduce psychology or our mental concepts to physical theories and physical concepts by general deterministic arguments is to try to reduce the rich facts of our mental activity to an unverifiable metaphysics of determinism. Kant had the story upside down: our mental life is empirically a rich phenomena which we can study scientifically and successfully. In contrast, the general theory of determinism as a view of the universe represents a metaphysics empty of content.

BIBLIOGRAPHY

Abelson, R.M. & Bradley, R.A. (1954). A 2 × 2 factorial with paired comparisons. *Biometrics*, **10**, 487–502.

Alekseev, V.M. (1968a). Quasirandom dynamical systems. I. Quasirandom diffeomorphisms. *Mathematicheskie USSR Sbornik 5*, 73–128.

Alekseev, V.M. (1968b). Quasirandom dynamical systems. II. One-dimensional nonlinear oscillations in a field with periodic perturbation. *Mathematicheskie USSR Sbornik 6*, 505–560.

Alekseev, V.M. (1969a). Quasirandom dynamical systems. III. Quasirandom oscillations of one-dimensional oscillators. *Mathematicheskie USSR Sbornik 7*, 1–43.

Alekseev, V.M. (1969b). Quasi-random dynamical systems. Doctoral Dissertation. *Mathematicheskie Zametki* **6**(4), 489–498. Translation in *Mathematical Notes, Academy of Sciences, USSR*, **6**(4), 749–753.

Angell, R.B. (1974). The geometry of visibles. *Noûs*, **8**, 87–117.

Anscombe, G.E.M. (1975). Causality and determination. In E. Sosa (Ed.), *Causation and Conditionals.* London, New York: Oxford University Press.

Archimedes. (1897). *The Works of Archimedes with the Method of Archimedes.* Translated by T.L. Heath. New York: Dover.

Aristotle, *On the Heavens.* Translated by W.K.G. Guthrie (1939). Cambridge, MA: Harvard University Press.

Aydelotte, W.O., Bogue, A.G. & Fogel, R.W. (Eds.). (1971). *The Dimensions of Quantitative Research in History.* Princeton, NJ: Princeton University Press.

Baianu, I.C. (1986). Computer models and automata theory in biology and medicine. *Mathematical Modeling*, **7**, 1513–1577.

Bartlett, F.C. (1932). *Remembering*. New York: Macmillan.

Basu, D. (1980). Randomization analysis of experimental data: The Fisher randomization test. *Journal of the American Statistical Association*, **75**, 575–595.

Battro, A.M., Netto, S.P., & Rozestraten, R.J.A. (1976). Riemannian geometries of variable curvature in visual space: Visual alleys, horopters, and triangles in big open fields. *Perception*, **5**, 9–23.

Bayes, T. (1763). An essay toward solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, **58**, 370–418.

Beale, H.K. (1946). What historians have said about the causes of the Civil War. In *Theory and practice in historical study: A report of the Committee on Historiography*. Social Science Research Council Bulletin, **54**, 53–102.

Becquerel, H. (1964). On the invisible radiations emitted by the salts of uranium. In A. Romer (Ed.) (1964), *The Discovery of Radioactivity and Transmutation*. New York: Dover.

Bell, J.S. (1964). On the Einstein-Podolsky-Rosen paradox. *Physics*, **1**, 195–200.

Bell, J.S. (1966). On the problem of hidden variables in quantum mechanics. *Reviews of Modern Physics*, **38**, 447–452.

Bell, J.S. (1971). Introduction to the hidden-variable question. In B. d'Espagnat (Ed.), *Foundations of quantum mechanics*, pp. 171–181. New York: Academic Press.

Berkeley G. (1901). An essay towards a new theory of vision. In A.C. Fraser (Ed.), *Berkeley's complete works: Vol. 1*. Oxford: Oxford University Press. (Original work published 1709)

Bickel, P.J., Hammel, E.A. & O'Connell, J.W. (1975). Sex bias in graduate admissions: Data from Berkeley. *Science*, **187**, 398–404.

Bjelke, B. (1975). Dietary vitamin A and human lung cancer. *International Journal of Cancer*, **15**, 561–565.

Blalock, H.M., Aganbegian, A., Borodkin, F.M., Boudon, R., & Capecchi, V. (Eds.). (1975). *Quantitative sociology: International Perspectives on Mathematical and Statistical Modeling*. New York: Academic Press.

Blank, A.A. (1953). The Luneburg theory of binocular visual space. *Journal of the Optical Society of America*, **43**, 717–727.

Blank, A.A. (1957). The geometry of vision. *British Journal of Physiological Optics*, **14**, 154–169.

Blank, A A. (1958a). Analysis of experiments in binocular space perception. *Journal of the Optical Society of America*, **48**, 911–925.

Blank, A. A. (1958b). Axiomatics of binocular vision: The foundations of metric geometry in relation to space perception. *Journal of the Optical Society of America*, **48**, 328–334.

Blank, A.A. (1961). Curvature of binocular visual space: An experiment. *Journal of the Optical Society of America*, **51**, 335–339.

Blumenfeld, W. (1913). Untersuchungen über die scheinbare Grösse in Sehraume. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, **65**, 241–404.

Bogolubov, N.N., Logunov, A.A., & Todorov, I.T. (1975). *Introduction to Axiomatic Quantum Field Theory* (English Trans.). Reading, MA: W.A. Benjamin, Inc.

Bouillier, F. (1868). *Histoire de la Philosophie Cartésienne*, (3rd Edition). Paris.

Bowerman, M. (1989). Learning a semantic system. What role do cognitive predispositions play? In M.L. Rice & R.L. Schiefelbusch (Eds.), *The Teachability of Language*, pp. 133–169. Baltimore: Paul H. Brookes.

Bradley, R.A. (1954a). Incomplete block rank analysis: On the appropriateness of the model for a method of paired comparisons. *Biometrics*, **10**, 375–390.

Bradley, R.A. (1954b). Rank analysis of incomplete block designs II. Additional tables for the method of paired comparisons. *Biometrika*, **41**, 502–537.

Bradley, R.A. (1955). Rank analysis of incomplete block designs III. Some large-sample results on estimation and power for a method of paired comparisons. *Biometrika*, **42**, 450–470.

Bradley, R.A. & Terry, M.E. (1952). Rank analysis of incomplete block designs I. The method of paired comparisons. *Biometrika*, **39**, 324–345.

Burks, A.W. (Ed.). (1984). *Essays on Cellular Automata.* Urbana, IL: University of Illinois Press.

Burks, C., & Farmer, D. (1984). Towards modelling DNA sequences as automata. In D. Farmer, T. Toffoli, & S. Wolfram (Eds.), *Cellular Automata* (Proceedings of an interdisciplinary workshop, Los Alamos, New Mexico, March 7–11, 1983), pp. 157–167. Amsterdam: North-Holland.

Busemann, H. (1955). *The Geometry of Geodesics.* New York: Academic Press.

Bush, R. & Mosteller, F. (1955). *Stochastic Models for Learning.* New York: Wiley.

Calvin, M. (1975). Chemical evolution. *American Scientist*, **63**, 169–177.

Campbell, N.R. (1920). *Physics: The elements.* Cambridge, MA: Cambridge University Press. Reprinted as *Foundations of Science: The philosophy of theory and experiment.* New York: Dover, 1957.

Campbell, N. R. (1928). *An Account of the Principles of Measurement and Calculation.* London: Longmans, Green.

Carden, T.S. (1974) The antibiotic problems (editorial). *The New Physician*, **19**.

Carnap R. (1938). Logical foundations of the unity of science. In O. Neurath, *et al.* (Eds.), *International Encyclopedia of Unified Science*, **1**, Part 1, pp. 42–62. Chicago, IL: University of Chicago Press

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

Chomsky, N. (1959). Review of B.F. Skinner. *Verbal Behavior Language*, **35**, 26–58.

Clauser, J.F., Horne, M.A., Shimony A. & Holt R.A. (1969). Proposed experiment to test local hidden-variable theories. *Physical Review Letter*, **23**, 880–884.

Codd, E. F. (1968). *Cellular Automata.* New York: Academic Press.

Cofer, C.N. (1973). Constructive processes in memory. *American Scientist*, **61**, 537–543.

Cohen, M.R., & Nagel, E. (1934). *An Introduction to Logic and Scientific Method*. New York: Harcourt, Brace & Co.

Coleman, J.S. (1964). *Introduction to Mathematical Sociology*. New York: Free Press.

Crangle, C. & Suppes, P. (1989a). Geometrical semantics for spatial prepositions. In P.A. French, T.E. Uehling, Jr., & H.K. Wettstein (Eds.), *Midwest Studies in Philosophy, Volume XIV*, pp. 399–422. Notre Dame, IN: Univ. of Notre Dame Press.

Crangle, C. & Suppes P. (1989b). Instruction dialogues: Teaching new skills to a robot. In G. Rodrigues (Ed.), *Proceedings of the NASA*

Conference on Space Telerobotics, (January 31–February 2, 1989, Pasadena, California). Pasadena, California: Jet Propulsion Laboratory, California Institute of Technology,

Crick, F., & Asanuma, C. (1986). Certain aspects of the anatomy and physiology of the cerebral cortex. In J.L. McClelland, D.E. Rumelhart, & the PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the microstructures of cognition. Vol. 2: Psychological and biological models*, pp. 333–371. Cambridge, MA, and London: Massachusetts Institute of Technology Press.

Cutting, J.E. (1986). *Perception with an Eye for Motion*. Cambridge, MA: The MIT Press.

Daniels, N. (1972). Thomas Reid's discovery of a non-euclidean geometry. *Philosophy of Science*, **39**, 219–234.

David, P.A., Gutman, H.G., Sutch, R., Temen, P., & Wright, G. (1976). *Reckoning with Slavery*. New York: Oxford University Press.

Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Oxford University Press.

Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

De Finetti, B. (1931). Sul significato della probabilità. *Fundamenta Mathematicae*, **17**, 298–329.

De Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de L'Institut Henri Poincaré*, **7**, 1–68. English translation in (H.E. Kyburg, Jr., & H.E. Smokler, (Eds.). (1964). *Studies in Subjective Probability*. New York: Wiley.

De Finetti, B. (1984). *Theory of Probability* (Vol. 1). New York: Wiley.

Dembowski, P. (1968). *Finite Geometries*. New York: Springer-Verlag.

Dempster, A.P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, **38**, 325–340.

Descartes, R. (1637). *La Geometrie: The geometry of Rene Descartes* (translated from the French and Latin by D.E. Smith & M.L. Latham). New York: Dover, 1954.

Descartes, R. *Oeuvres de Descartes*, VIII. Adam and Tannery (Eds.) (1897). Paris: J. Vrin.

Diaconis, P. (1977). Finite forms of de Finetti's theorem on exchangeability. *Synthese*, **36**, 271–281.

Dijksterhuis, E.J. (1956). *Archimedes*. New York: Humanities Press.

Eells, E. & Sober, E. (1983). Probabilistic causality and the question of transitivity. *Philosophy of Science*, **50**, 33–57.

Ehrenfeucht, A., & Mycielski, J. (1973a). Interpolation of functions over a measure space and conjectures about memory. *Journal of Approximation Theory*, **9**, 218–236.

Ehrenfeucht, A., & Mycielski, J. (1973b). Organization of memory. *Proceedings of the National Academy of Sciences U.S.A.*, **70**, 1478–1480.

Ehrenfeucht, A., & Mycielski, J. (1977). Learnable functions. In R.E. Butts & J. Hintikka (Eds.), *Foundational Problems in the Special Sciences*, (Proceedings of the Fifth International Congress of Logic, Methodology and Philosophy of Science, London, Ontario, Canada), Vol. 2, pp. 251–256. Dordrecht, Holland: Reidel.

Estes, W.K. (1950). Toward a statistical theory of learning. *Psychological Review*, **57**, 94–107.

Estes, W.K. (1959). Component and pattern models with Markovian interpretations. In R.R. Bush & W.K. Estes (Eds.), *Studies in Mathematical Learning Theory*. Stanford: Stanford University Press.

Estes, W.K. & Suppes, P. (1959a). Foundations of linear models. In R.R. Bush & W.K. Estes (Eds.), *Studies in Mathematical Learning Theory*, pp. 137–179. Stanford, CA: Stanford University Press.

Estes, W. K. & Suppes, P. (1959b). *Foundations of Statistical Learning Theory, II: The stimulus sampling model*. (Technical Report # 26). Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, CA.

Estes, W.K. & Suppes, P. (1974). Foundations of stimulus sampling theory. In D.H. Krantz, R.C. Atkinson, R.D. Luce, & P. Suppes (Eds.), *Contemporary Developments in Mathematical Psychology, Vol. 1: Learning, memory and thinking*. San Francisco, CA: Freeman.

Euclid. (1945). Optics. (H.E. Burton, Trans.). *Journal of the Optical Society of America*, **35**, 357–372.

Fararo, T.J. (1973). *Mathematical Sociology: An introduction to fundamentals*. New York: Wiley.

Fine, A. (1982). Hidden variables, joint probability, and the Bell inequalities. *Physical Review Letter*, **48**, 291–295.

Fogel, R.E. & Engerman, S.L. (1974). *Time on the Cross: The economics of American negro slavery*. Boston, MA: Little, Brown.

Foley, J.M. (1964). Desarguesian property in visual space. *Journal of the Optical Society of America*, **54**, 684–692.

Foley, J.M. (1965). Visual space: A scale of perceived relative direction. *Proceedings of the 73rd Annual Convention of the American Psychological Association*, **1**, 49–50.

Foley, J.M. (1966). Locus of perceived equidistance as a function of viewing distance. *Journal of the Optical Society of America*, **56**, 822–827.

Foley, J.M. (1969). Distance in stereoscopic vision: The three-point problem. *Vision Research*, **9**, 1505–1521.

Foley, J.M. (1972). The size-distance relation and intrinsic geometry of visual space: Implications for processing. *Vision Research*, **12**, 323–332.

Foley, J.M. (1978). Distance perception. In R. Held, H. Leibowitz, & H.L. Teuber (Eds.), *Handbook of Sensory Physiology, Vol. 8: Perception*. New York, Berlin: Springer-Verlag.

Ford, L. R., Jr. (1957). Solution of a ranking problem from binary comparisons. *American Mathematical Monthly*, Herbert Ellsworth Slaught Memorial Papers, **64**, 28–33.

Freudenthal, H. (1965). Lie groups in the foundations of geometry. *Advances in Mathematics*, **1**, 145–190.

Friedman, M. (1957). *A Theory of the Consumption Function*. National Bureau of Economic Research, New York.

Fu, K.S. (1974). *Syntactic Methods in Pattern Recognition*. New York: Academic Press.

Gale, D. (1952). An indeterminate problem in classical mechanics. *American Mathematical Monthly*, **59**, 291–295.

Genovese, E.D. (1974). *Roll, Jordan, Roll: The world the slaves made*. New York: Pantheon.

Giere, R.N. (1973). Objective single-case probabilities and the foundations of statistics. In P. Suppes, L. Henkin, G. C. Moisil & A. Joja (Eds.), *Logic, Methodology, and Philosophy of Science, IV*, pp. 467–483. Amsterdam: North-Holland.

Gogel, W.C. (1956a). Relative visual direction as a factor in relative distance perceptions. *Psychological Monographs*, **70**, (11, No. 418).

Gogel, W.C. (1956b). The tendency to see objects as equidistant and its inverse relation to lateral separation. *Psychological Monographs*, **70**, (4, No. 411).

Gogel, W.C. (1963). The visual perception of size and distance. *Vision Research*, **3**, 101–120.

Gogel, W.C. (1964a). The perception of depth from binocular disparity. *Journal of Experimental Psychology*, **67**, 379–386.

Gogel, W.C. (1964b). Size cue to visually perceived distance. *Psychological Bulletin*, **62**, 217–235.

Gogel, W.C. (1965). Equidistance tendency and its consequences. *Psychological Bulletin*, **64**, 153–163.

Gold, E.M. (1967). Language identification in the limit. *Information and Control*, **10**, 447–474.

Good, I.J. (1961/1962). A causal calculus. *British Journal of Science*, **11**, 305–318; **12**, 43–51; **13**, 88.

Good, I.J. (1962). Subjective probability as the measure of a non-measurable set. In E. Nagel, P. Suppes & A. Tarski (Eds.), *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*. Stanford, CA: Stanford University Press.

Gorry, G.A., & Barnett, G.O. (1967–1968). Experience with a model of sequential diagnosis. *Computers and Biomedical Research*, **1**, 490–507.

Granger, C.W.J. (1969). Investigating causal relations by econometric models and cost-spectral methods. *Econometrica*, **37**, 424–438.

Greenwood, M., & Yule, G.U. (1915). The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general. *Proceedings of the Royal Society of Medicine*, **8**, 113–190.

Grossberg, S. (1974). Classical and instrumental learning by neural networks. In R. Rosen & F. Snell (Eds.), *Progress in Theoretical Biology*. New York: Academic Press.

Grossberg, S. (1978). A theory of visual coding, memory, and development. In E.L.J. Leeuwenberg & H.F.J.M. Buffart (Eds.), *Formal Theories of Visual Perception*, pp. 7–26. New York: Wiley.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, **87**, 1–51.

Grossberg, S. (1982). Associative and competitive principles of learning and development: The temporal unfolding and stability of STM and LTM patterns. In S. Aman & M.A. Arbib (Eds.), *Competition and Cooperation in Neural Nets: Lecture notes in biomathematics*, **45**, 295–341.

Grossberg, S., & Kuperstein, M. (1986). *Neural Dynamics of Adaptive Sensory-Motor Control*. Amsterdam: North-Holland.

Grünbaum, A. (1963). *Philosophical Problems of Space and Time*. New York: Knopf.

Grzegorczyk, A. (1955). The Systems of Lesniewski in relation to contemporary logical research. *Studia Logica*, **3**, 77–95.

Hacking, I. (1975). *The Emergence of Probability*. London: Cambridge University Press.

Hall, R.E. & Mishkin, F.S. (1982). The sensitivity of consumption to transitory income: Estimates from panel data on households. *Econometrica*, **50**, 461–481.

Hamburger, H., & Wexler, K. (1973). Identifiability of a class of transformational grammars. In K.J.J. Hintikka, J.M.E. Moravcsik, & P. Suppes (Eds.), *Approaches to Natural Language*, pp. 153–166. Dordrecht, Holland: D. Reidel.

Hamburger, H., & Wexler, K. (1975). A mathematical theory of learning transformational grammar. *Journal of Mathematical Psychology*, **12,** 137–177.

Hardy, L.H., Rand, G., Rittler, M.C., Blank, A.A., & Boeder, P. (1953). *The Geometry of Binocular Space Perception*. Elizabeth, NJ: Schiller.

Harrison, M.A. (1965). *Introduction to Switching and Automata Theory*. New York: McGraw-Hill.

Hawking, S.W. & Ellis, G.F.R. (1973). *The Large Scale Structure of Space-Time*. Cambridge, MA: Cambridge University Press.

Heath, T.L. (Ed.). (1897). *The Works of Archimedes with the Method of Archimedes*. New York: Dover.

Hegel, G.W.F. (1899). *The Philosophy of History* (J. Sibree, Trans.). Colonial Press. Reprinted by Dover Publications, New York, 1956.

Helmholtz, H. (1868). Über die Tatsachen, die der Geometrie zugrunde liegen. *Wissenschaftliche Abhandlungen*, **2**, 618–637.

Hesslow, G. (1976). Two notes on the probabilistic approach to causality. *Philosophy of Science*, **43**, 290–292.

Hesslow, G. (1981). Causality and determinism. *Philosophy of Science*, **48**, 591–605.

Hillenbrand, F. (1902). Theorie der scheinbaren Grösse bei binocularem Sehen. *Denkschriften* (Akademie der Wissenschaften in Wien, Mathematisch-Naturwissenschaftliche Klasse), **7**, 255–307.

Hintzman, D.L. (1980). Simpson's paradox and the analysis of memory retrieval. *Psychological Review*, **87**, 398–410.

Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems.* Ann Arbor, MI: University of Michigan Press.

Hopf, E. (1934). On causality, statistics and probability. *Journal of Mathematics and Physics*, **17**, 51–102.

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, **79**, 2554–2558.

Hopfield, J.J., & Tank, D.W. (1985). "Neural" computation of decisions in optimization problems. *Biological Cybernetics*, **52**, 141–152.

Horgan, T. (1980). Humean causation and Kim's theory of events. *Canadian Journal of Philosophy*, **10**, 663–679.

Hosoya, Y. (1977). On the Granger condition for non-causality. *Econometrica*, **7**, 1735–1736.

Hull, C.L. (1943). *Principles of Behavior.* New York: Appleton.

Hull, C.L., Hovland, C.I., Ross, R.T., Hall, M., Perkins, D.T., & Fitch, F.B. (1940). *Mathematico-deductive Theory of Rote Learning.* New Haven, CT: Yale University Press.

Hume, D. (1879). *The History of England from the Invasion of Julius Caesar to the Abdication of James the Second*, 1688 (Vol. V). New York: Harper.

Humphreys, P. (1976). *Inquiries in the Philosophy of Probability: Randomness and independence.* Unpublished Ph.D. Dissertation, Stanford University, Xerox University Microfilms Publication No. 76–18774.

Indow, T. (1967). Two interpretations of binocular visual space: Hyperbolic and Euclidean. *Annals of the Japan Association for Philosophy of Science*, **3**, 51–64.

Indow, T. (1968). Multidimensional mapping of visual space with real and simulated stars. *Perception and Psychophysics*, **3**, 45–53.

Indow, T. (1974a). Applications of multidimensional scaling in perception. In E.C. Carterette & M.P. Friedman (Eds.), *Handbook of Perception, Vol. 2: Psychophysical judgment and measurement*, pp. 493–531. New York: Academic Press.

Indow, T. (1974b). On geometry of frameless binocular perceptual space. *Psychologia*, **17**, 50–63.

Indow, T. (1975). An application of MDS to study of binocular visual space. In *Theory, Methods and Applications of Multidimensional Scaling and Related Techniques*, U.S.-Japan Seminar (sponsored by the

National Science Foundation and Japan Society for the Promotion of Science). University of California, San Diego.

Indow, T. (1979). Alleys in visual space. *Journal of Mathematical Psychology*, **19**, 221–258.

Indow, T. (1982). An approach to geometry of visual space with no a priori mapping functions: Multidimensional mapping according to Riemannian metrics. *Journal of Mathematical Psychology*, **26**, 204–236.

Indow, T., Inoue, E., & Matsushima, K. (1962a). An experimental study of the Luneburg theory of binocular space perception (1): The 3- and 4-point experiments. *Japanese Psychological Research*, **4**(1), 6–16.

Indow, T., Inoue, E., & Matsushima, K. (1962b). An experimental study of the Luneburg theory of binocular space perception (2): The alley experiments. *Japanese Psychological Research*, **4**(1), 17–24.

Indow, T., Inoue, E., & Matsushima, K. (1963). An experimental study of the Luneburg theory of binocular space perception (3): The experiments in a spacious field. *Japanese Psychological Research*, **5**(1), 10–27.

Jamison, D., Lhamon, D., & Suppes, P. (1970). Learning and the structure of information. In J. Hintikka & P. Suppes (Eds.), *Information and Inference*. Holland: Reidel.

Jammer (1957). *Concepts of Force: A study in the foundations of dynamics*. Cambridge, MA: Harvard University Press.

Kahn, L. (1918). *Metaphysics of the Supernatural as Illustrated by Descartes*. New York: Columbia Univ. Press.

Kandel, E.R. (1985). Cellular mechanisms of learning and the biological basis of individuality. In E.R. Kandel & J.H. Schwartz (Eds.), *Principles of Neural Science* (2nd ed.). Amsterdam: Elsevier.

Kanerva, P. (1988). *Space Distributed Memory.* Cambridge, MA: Massachusetts Institute of Technology Press.

Kant, I. (1786). *Die Metaphysischen Anfangsgründe der Naturwissenschaft.* (Reprinted as *Metaphysical Foundations of Natural Science.* J. Ellington (Trans.). Indianapolis: Bobbs-Merrill, 1970.

Kaplan, R.M., & Bresnan, J. (1982). Lexical-functional grammar: A formal system for grammatical representation. In J. Bresnan (Ed.), *The Mental Representation of Grammatical Relations.* Cambridge, MA: Massachusetts Institute of Technology Press.

Kauffman, S.A. (1984). Emergent properties in random complex automata. In D. Farmer, T. Toffoli, & S. Wolfram (Eds.), *Cellular automata* (Proceedings of an Interdisciplinary Workshop, Los Alamos, New Mexico, March 7–11, 1983), pp. 145–156. Amsterdam: North-Holland.

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

Kendall, M.G., & Stuart, A. (1961). *The Advanced Theory of Statistics, Vol. 2: Inference and Relationship.* London: Griffin & Co.

Klein, F. (1893). A comparative review of recent researches in geometry. *Bulletin of the New York Mathematical Society*, **2**, 215–249.

Kolmogorov, A. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Berlin: J. Springer. English edition, New York: Chelsea, 1950.

Kolgomorov, A.N. (1950b). *Foundations of the Theory of Probability.* New York: Chelsea.

Koopman, B.O. (1940a). The axioms and algebra of intuitive probability. *Annals of Mathematics*, **41**, 269–292.

Koopman, B.O. (1940b). The bases of probability. *Bulletin of the American Mathematical Society*, **46**, 763–774.

Kraft, C.H., Pratt, J.W. & Seidenberg, A. (1959). Intuitive probability on finite sets. *Annals of Mathematical Statistics*, **30**, 408–419.

Krantz, D.H., Luce, R.D., Suppes, P. & Tversky A. (1971). *Foundations of Measurement, Vol I.* New York: Academic Press.

Kreisel, G. (1967). Informal rigour and completeness proofs. In I. Lakatos (Ed.), *Problems in the Philosophy of Mathematics*, (Vol. 1 of the Proceedings of the International Colloquium in the Philosophy of Science), pp. 138–171. Amsterdam: North-Holland.

Kunin, C.M., Tupasi, T., & Craig, W.A. (1973). Use of antibiotics: A brief exposition of the problem and some tentative solutions. *Annals of Internal Medicine*, **79**, 555–560.

Lamb, H. (1919). The kinematics of the eye. *Philosophical Magazine*, **38**, 685–695.

Langton, C.G. (1984). Self-reproduction in cellular automata. *Physica*, **10D**, 135–144. Reprinted in D. Farmer, T. Toffoli, & S. Wolfram (Eds.), *Cellular Automata* (Proceedings of an Interdisciplinary Workshop, Los Alamos, New Mexico, March 7–11, 1983), pp. 135–144. Amsterdam: North-Holland.

Levelt, W.J.M. (1982). Cognitive styles in the use of spatial direction terms. In R.J. Jarvella & W. Klein (Eds.), *Speech, Place, and Action*, pp. 251–268. Chichester, New York: John Wiley & Sons Ltd.

Levelt, W.J.M. (1984). Some perceptual limitations on talking about space. In A.J. Van Doorn, W.A. van de Grind, & J. Koenderink (Eds.), *Limits in Perception*, pp. 323–358. Ultrecht: VNU Press.

Lindberg, D.C. (1970). *John Pecham and the Science of Optics: Perspectiva communis*. Madison, WI: University of Wisconsin Press.

Luce, R.D. (1956). Semiorders and a theory of utility discrimination. *Econometrica*, **24**, 178–191.

Luce, R.D. (1959). *Individual Choice Behavior*. New York: Wiley.

Luce, R.D. (1967). Sufficient conditions for the existence of a finitely additive probability measure. *Annals of Mathematical Statistics*, **30**, 408–419.

Luneberg, R.K. (1947). *Mathematical Analysis of Binocular Vision*. Princeton, NJ: Princeton University Press.

Luneberg, R.K. (1948). Metric methods in binocular visual perception. In *Studies and Essays* (Courant Anniversary Volume), pp. 215–240. New York: Interscience.

Luneberg, R.K. (1950). The metric of binocular visual space. *Journal of the Optical Society of America*, **40**, 627–642.

Mach, E. (1942). *The Science of Mechanics* (5th English ed.; T.J. McCormack, Trans.). La Salle, IL: Open Court.

Mackey, G. W. (1957). Quantum mechanics and Hilbert space. *American Mathematical Monthly*, **64S2**, 45–57.

Mackey, G. W. (1963). *The Mathematical Foundations of Quantum Mechanics*. New York: Benjamin.

Malinvaud, E. (1966). *Statistical Methods of Econometrics*. Amsterdam: North Holland Press.

Martin, E. (1981). Simpson's paradox resolved: A reply to Hintzman. *Psychological Review*, **88**, 372–374.

Martin, R.M. (1981). On the language of causal talk: Scriver and Suppes. *Theory and Decision*, **13**, 331–344.

Matsushima, K. & Noguchi, H. (1967). Multidimensional representation of binocular visual space. *Japanese Psychological Research*, **9**, 85–94.

McCulloch, W.S., & Pitts, W. (1943). A logical calculus of the ideas imminent in nervous activity. *Bulletin of Mathematical Biophysics*, **5**, 115–133.

McKeon, R. (Ed.). (1941). *The Basic Works of Aristotle.* New York: Random House.

McKeown, T. (1976). *The Modern Rise in Population.* London: Edward Arnold.

Mellor, D.H. (1971). *The Matter of Chance.* Cambridge, England: Cambridge University Press.

Minsky, M., & Papert, S. (1969). *Perceptrons.* Cambridge, MA: Massachusetts Institute of Technology Press.

Moler, N. & Suppes, P. (1968). Quantifier-free axioms for constructive plane geometry. *Composito Mathematica*, **20**, 143–152.

Moser, J. (1973). *Stable and Random Motions in Dynamical Systems with Special Emphasis on Celestial Mechanics.* Hermann Weyl Lectures, The Institute for Advanced Study. Princeton, NJ: Princeton University Press.

Mouy, P. (1934). *Le Développement de la Physique Cartésienne* 1646–1712. Paris: Librairie Philosophique J. Vrin.

Moyal, J.E. (1949). Quantum mechanics as a statistical theory. *Proceedings of the Cambridge Philosophical Society*, **45**, 99–124.

Moser, J. (1973). *Stable and Random Motions in Dynamical Systems with Special Emphasis on Celestial Mechanics.* Herman Weyl Lectures, The Institute for Advanced Study. Princeton, NJ: Princeton University Press.

Myhill, J. (1970). The abstract theory of self-reproduction. In A.W. Burks (Ed.), *Essays on Cellular Automata*, pp. 206–218. Urbana, IL: University of Illinois Press.

Nelson, E. (1966). Derivation of the Schrödinger equation from Newtonian mechanics. *Physical Review*, **150**, 1079–1085.

Nelson, E. (1967). *Dynamical Theories of Brownian Motion.* Princeton, NJ: Princeton University Press.

Nelson, E. (1985). *Quantum Fluctuations.* Princeton, NJ: Princeton University Press.

Nelson, R.J. (1975). Behaviorism, finite automata, and stimulus-response theory. *Theory and Decision*, **6**, 249.

Neugebauer, O. (1957). *The Exact Sciences in Antiquity* (2nd ed.). Providence, RI: Brown University Press.

Neugebauer, O. (1975). *A History of Ancient Mathematical Astronomy*, (3 Volumes). New York: Springer-Verlag.

Neurath, O. (1938). Unified science as encyclopedic integration. In O. Neurath, *et al.* (Eds.), *International Encyclopedia of United Science*, Volume 1, Part 1, pp. 1–27. Chicago: University of Chicago Press.

Newell, A. (1980). Physical symbol systems. *Cognitive Science*, **4**, 135–183.

Newell, A. & Simon, H.A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Newton, I. (1931). *Opticks* (Reprinted from the 4th ed.). London: Bell. (Original work published 1704)

Newton, I. (1946). *Philosophiae Naturalis Principia Mathematica*, (Cajori Trans.). Berkeley, CA: University of California Press.

Nishikawa, Y. (1967). Euclidean interpretation of binocular visual space. *Japanese Psychological Research*, **9**, 191–198.

Noll, W. (1959). The foundations of classical mechanics in the light of recent advances in continuum mechanics. In L. Henkin, P. Suppes, & A. Tarski (Eds.), *The Axiomatic Method*, pp. 266–281. Amsterdam: North-Holland.

Noll, W. (1966). The foundations of mechanics. In C. Truesdell & G. Grioli (coordinators), *Non-linear Continuum Theories*, pp. 159–200. Rome: Edizioni Cremonese.

Norman, M.F. (1972). *Markov Processes and Learning Models*. New York: Academic Press.

Oakes, W.J. (Ed.). (1940). *Stoic and Epicurean Philosophers: The complete extant writings of Epicureus, Epictetus, Lucretius, and Marcus Aurelius*. New York: Random House.

Osherson, D.N., Stob, M., & Weinstein, S. (1986). *Systems that Learn*. Cambridge, MA: Massachusetts Institute of Technology Press.

Peterson, O.L., Andrews, L.P., Spain, R.S., & Greenberg, B.G. (1956). An analytical study of North Carolina general practice. *Journal of Medical Education*, **31**, 1–165.

Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Pinker, S. & Mehler J. (Eds.). (1988). *Connections and Symbols.* Cambridge, MA: The MIT Press.

Poincaré, H. (1912). *Calcul des probabilités* (2nd ed.). Paris: Gauthier-Villars.

Poincaré, H. (1913). *Science and Hypothesis.* Lancaster, PA: The Science Press.

Popper, K.R. (1957). The propensity interpretation of the calculus of probability and the quantum theory. In S. Körner (Ed.), *Observation and Interpretation in the Philosophy of Physics*, pp. 65–70. London: Butterworth.

Popper, K.R. (1959). The propensity interpretation of probability. *The British Journal for the Philosophy of Science*, **10**(37), 25–42.

Popper, K.R. (1967). Quantum mechanics without 'The Observer'. In M. Bunge (Ed.), *Quantum Theory and Reality*, pp. 7–44. Berlin: Springer-Verlag.

Popper, K.R. (1968). Birkhoff and von Neumann's interpretation of quantum mechanics. *Nature*, **219**, 682–685.

Popper, K.R. (1974). Replies to my critics. In P.A. Schilpp (Ed.), *The Philosophy of Karl Popper* (Vol. 2). La Salle, IL: Open Court.

Price, D.J. de S. (1961). *Science since Babylon.* New Haven, CT: Yale University Press.

Ptolemy, C. (1951). *The Almagest.* Translated by R.C. Taliaferro. In *Great Books of the Western World*, Vol. 21. Chicago: Encyclopedia Britannica.

Ptolemy, C. (1984). *The Almagest.* Translated by G. J. Toomer. New York: Springer Verlag.

Ramsey, F.P. (1950). *The Foundations of Mathematics and other Logical Essays.* Edited by R.B. Braithwaite, New York: Humanities Press.

Reid, T. (1967). Inquiry into the human mind. In *Philosophical Works of Thomas Reid, Vol. 1: George Olms.* Hildesheim, Germany: Verlag's Buchhandlung. (Original work published 1764)

Riemann, B. (1866/1867). Über die Hypothesen, welche der Geometrie zu Grunde liegen. *Gesellschaft der Wissenschaften zu Göttingen: Abhandlungen*, **13**, 133–152.

Robb, A.A. (1936). *Geometry of Space and Time.* Cambridge, MA: Cambridge University Press.

Roberts, A.W., & Visconti, J.A. (1972). The rational and irrational use of systemic antimicrobial drugs. *American Journal of Hospital Pharmacy*, **29**, 828–834.

Roberts, F.S. (1970). On nontransitive indifference. *Journal of Mathematical Psychology*, **7**, 243–258.

Roberts, F.S. (1973). Tolerance geometries. *Notre Dame Journal of Formal Logic*, **14**, 68–76.

Roberts, F.S. & Suppes, P. (1967). Some problems in the geometry of visual perception. *Synthese*, **17**, 173–201.

Rosen, D. (1978). In defense of a probabilistic theory of causality. *Philosophy of Science*, **45**, 604–613.

Rosen, D. (1982). A critique of deterministic causality. *Philosophical Forum*, **14**(2), 101–130.

Rosenblatt, F. (1959). *Two Theorems of Statistical Separability in the Perceptron* (Proceedings of a symposium on the mechanization of thought processes). London: HM Stationery Office.

Rumelhardt, D.E., McClelland, J.L., & the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the microstructures of cognition* (2 vols.). Cambridge, MA: Massachusetts Institute of Technology Press.

Rutherford, E. (1964). A radio-active substance emitted from thorium compounds. In A. Romer (Ed.) (1964), *The Discovery of Radioactivity and Transmutation*. New York: Dover.

Salmon, W.C. (1980). Probabilistic causality. *Pacific Philosophical Quarterly*, **61**, 50–74.

Salmon, W.C. (1982). Further reflections. In R. McLaughlin (Ed.), *What? Where? When? Why?*, pp. 231–280. Dordrecht, Holland: D. Reidel Publishing Co.

Savage, L.J. (1954). *The Foundations of Statistics*. New York: Wiley.

Scheckler, W.E., & Bennett, J.V. (1970). Antibiotic usage in seven community hospitals. *Journal of the American Medical Association*, **213**, 264–267.

Schelling, H. (1956). Concept of distance in affine geometry and its applications in theories of vision. *Journal of the Optical Society of America*, **46**, 309–315.

Schmidt, O. (1975). A system of axioms for the Archimedean theory of equilibrium and center of gravity. *Centaurus*, **19**, 2–35.

Scott, D. (1964). Measurement models and linear inequalities. *Journal of Mathematical Psychology*, **1**, 233–247.

Scott, D., & Suppes, P. (1958). Foundational aspects of theories of measurement. *Journal of Symbolic Logic*, **23**, 113–128.

Shepherdson, J.C., & Sturgis, H.E. (1963). The computability of partial recursive functions. *Journal of the Association of Computing Machinery*, **10**, 217–255.

Shortliffe, E.H. (1976). *Computer-based Medical Consultations: MYCIN*. New York: American Elsevier Publishing Co.

Simmons, H.E., & Stolley, P.D. (1974). This is medical progress?—Trends and consequences of antibiotic use in the United States. *Journal of the American Medical Association*, **227**, 1023–1026.

Simpson, E.H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society* (Ser. B), **13**, 238–241.

Sims, C.A. (1972). Money, income, and causality. *The American Economic Review*, **62**, 540–552.

Sitnikov, K. (1960). Existence of oscillating motions for the three-body problem. *Doklady Akademii Nauk, USSR*, **133**(2), 303–306.

Skinner, B.F. (1977). *Verbal Behavior*. New York: Appleton.

Smith, C.A.B. (1961). Consistency in statistical inference and decision. *Journal of the Royal Statistical Society* B, **23**, 1–25.

Smith, B. H., & Kreutzberg, G.W. (1976). Neuron-target cell interactions. *Neurosciences Research Program Bulletin*, **14**, 211–453.

Smoluchowski, M. (1918). Über den Begriff des Zufalls und den Ursprung der Wahrscheinlichkeitsgesetze in der Physik. *Die Naturwissenschaften*, **17**, 253–263.

Stein, W. (1930). Der Begriff des Schwerpunktes bei Archimedes. Quellen und Studien zur Geschichte der Mathematik. *Physik und Astronomie*, **1**, 221–244.

Stock, H. (1931). *The Method of Descartes in the Natural Sciences*. Jamaica, NY: The Marion Press.

Stoll, E.A. (1962). *Geometrical concept formation in kindergarten children*. Ph.D. Thesis, Stanford University, Stanford, CA.

Strawson, P.F. (1966). *The Bounds of Sense: An essay on Kant's Critique of Pure Reason*. London: Methuen.

Suppes, P. (1951). A set of independent axioms for extensive quantities. *Portugaliae Mathematica*, **10**, 163–172.

Suppes, P. (1956). The role of subjective probability and utility in decision-making. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium of Mathematical Statistics and Probability* (1954–1955, Vol. 5). Berkeley, CA: University of California Press.

Suppes, P. (1957) *Introduction to Logic*. New York: Van Nostrand.

Suppes, P. (1959a). A linear model for a continuum of responses. In R.R. Bush & W.K. Estes (Eds.), *Studies in Mathematical Learning Theory*, pp. 400–414. Stanford, CA: Stanford University Press.

Suppes, P. (1959b). Axioms for relativistic kinetics with or without parity. In L. Henkin, P. Suppes, & A. Tarski (Eds.), *The Axiomatic Method with Special Reference to Geometry and Physics*. (Proceedings of an international symposium held at the University of California, Berkeley, December 16, 1957 - January 4, 1958), pp. 291–307. Amsterdam: North-Holland.

Suppes, P. (1961). Probability concepts in quantum mechanics. *Philosophy of Science*, **28**(4), 378–389.

Suppes, P. (1962). Models of data. In E. Nagel, P. Suppes & A. Tarski (Eds.), *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 international congress*, pp. 252–261. Stanford, CA: Stanford University Press.

Suppes, P. (1965). On the behavioral foundations of mathematical concepts. *Monographs of the Society for Research in Child Development*, **30**, 60–96. Reprinted in Suppes (1969a).

Suppes, P. (1966). Probabilistic inference and the concept of total evidence. In J. Hintikka & P. Suppes (Eds.), *Aspects of Inductive Logic*, pp. 49–65. Amsterdam: North-Holland.

Suppes, P. (1967). Some extensions of Randall's interpretation of Kant's philosophy of science. In J.P. Anton (Ed.), *Naturalism and Historical Understanding: Essays on the philosophy of John Herman Randall, Jr.*, pp. 108–120. New York: State Univ. of N.Y. Press.

Suppes, P. (1968). The desirability of formalization in science. *Journal of Philosophy*, **65**, 651–664.

Suppes, P. (1969a). *Studies in the Methodology and Foundations of Science*. Dordrecht, Holland: D. Reidel Publishing Co.

Suppes, P. (1969b). Stimulus-response theory of finite automata. *Journal of Mathematical Psychology*, **6**, 327–355.

Suppes, P. (1970). *A Probabilistic Theory of Causality*. Amsterdam: North-Holland.

Suppes, P. (1972). Some open problems in the philosophy of space and time. *Synthese*, **24**, 298–316.

Suppes, P. (1973a). Facts and fantasies of education. In M.C. Wittrock (Ed.), *Changing Education: Alternatives from educational research*, pp. 6–45. Englewood Cliffs, NJ: Prentice-Hall.

Suppes, P. (1973b). New foundations of objective probability: Axioms of propensities. In P. Suppes, L. Henkin, G.C. Moisil, & A. Joja (Eds.), *Logic, Methodology, and Philosophy of Science IV: Proceedings of the Fourth International Congress for Logic, Methodology and Philosophy of Science, Bucharest, 1971*, pp. 515–529. Amsterdam: North-Holland.

Suppes, P. (1974a). Aristotle's concept of matter and its relation to modern concepts of matter. *Synthese*, **28**, 27–50.

Suppes, P. (1974b). The measurement of belief. *Journal of the Royal Statistical Society* (Series B), **36**, 160–175.

Suppes, P. (1974c). Popper's analysis of probability in quantum mechanics. In P.A. Schilpp (Ed.), *The Philosophy of Karl Popper*, Vol. 2, pp. 760–774. La Salle, IL: Open Court. [Article this volume]

Suppes, P. (1974d). The essential but implicit role of modal concepts in science. In K.F. Schafner & R.S. Cohen (Eds.), *Philosophy of Science Association, 1972: Proceedings*, pp. 305–314. Dordrecht, Holland: D. Reidel.

Suppes, P. (1975). From behaviorism to neobehaviorism. *Theory and Decision*, **6**, 269–285. [Article this volume]

Suppes, P. (1977a). A survey of contemporary learning theories. In R.E. Butts & J. Hintikka (Eds.), *Foundational Problems in the Special Sciences* (Part two of the Proceedings of the Fifth International Congress of Logic, Methodology and Philosophy of Science, London, Ontario, Canada, 1975), pp. 217–239. Dordrecht, Holland: D. Reidel.

Suppes, P. (1977b). Learning theory for probabilistic automata and register machines. In H. Spada & W. F. Kempf (Eds.), *Structural Models of Thinking and Learning* (Proceedings of the 7th IPN-Symposium on Formalized Theories of Thinking and Learning and their Implications for Science Instruction), pp. 57–79. Bern: Hans Huber Publishers.

Suppes, P. (1980). Procedural semantics. In R. Haller & W. Grassl (Eds.), *Proceedings of the Fourth International Wittgenstein Symposium, August 28 - September 2, 1979, Kirchberg, Austria*, pp. 27–35. Vienna: Hölder-Pichler-Tempsky.

Suppes, P. (1981). The plurality of science. In P. Asquith & I. Hacking (Eds.), *PSA 1978* (Vol. 2). East Lansing, MI: Philosophy of Science Association.

Suppes, P. (1982). Variable-free semantics with remarks on procedural extensions. In T.W. Simon & R.J. Scholes (Eds.), *Language, Mind, and Brain*. Hillsdale, NJ: Lawrence Erlbaum.

Suppes, P. (1983). Learning language in the limit, review of K. Wexler & P.W. Culicover, *Formal Principles of Language Acquisition*. *Contemporary Psychology*, **28**, 5–6.

Suppes, P. (1984). *Probabilistic Metaphysics*. Oxford: Blackwell.

Suppes, P. (1987). Propensity representations of probability. *Erkenntnis*, **26**, 335–358.

Suppes, P. (1988). Comment: Causality, complexity and determinism. *Statistical Science*, **3**, 398–403.

Suppes, P. (1990). Probabilistic causality in quantum mechanics. *Journal of Statistical Planning and Inference*, **25**, 293–302.

Suppes, P. (1991a). Indeterminism or instability, does it matter? In G.G. Brittan, Jr. (Ed.), *Causality, Method, and Modality*, pp. 5–22. Dordrecht, Netherlands: Kluwer.

Suppes, P. (1991b). *Language for Humans and Robots*. Oxford: Blackwell.

Suppes, P. & Atkinson, R.C. (1960). *Markov Learning Models for Multiperson Interactions*. Stanford, CA: Stanford University Press.

Suppes P., Fletcher, J.D., & Zanotti, M. (1976). Models of individual trajectories in computer-assisted instruction for deaf students. *Journal of Educational Psychology*, **68**, 117–127.

Suppes, P. & Ginsberg, R. (1963). A fundamental property of all-or-none models, binomial distribution of responses prior to conditioning, with application to concept formation in children. *Psychological Review*, **70**, 139–161.

Suppes, P., Krantz, D.H., Luce, R.D., & Tversky, A. (1989). *Foundations of Measurement, Volume II: Geometrical, threshold, and probabilistic representations*. San Diego, CA: Academic Press, Inc.

Suppes, P., Macken, E., & Zanotti, M. (1978). The role of global psychological models in instructional technology. In R. Glaser (Ed.), *Advances in Instructional Psychology, Vol. 1*, pp. 229–259. Hillsdale, NJ: Erlbaum.

Suppes, P., & Morningstar, M. (1972). *Computer-assisted Instruction at Stanford, 1966–68: Data, models, and evaluation of the arithmetic programs*. New York: Academic Press.

Suppes, P., & Rottmayer, W. (1974). Automata. In E.C. Carterette, & M.P. Friedman (Eds.), *Handbook of Perception, Vol. 1: Historical and philosophical roots of perception*, pp. 335–362. New York: Academic Press.

Suppes, P., Rouanet, H., Levine, M., & Frankmann, R.W. (1964). Empirical comparison of models for a continuum of responses with noncontingent bimodal reinforcement. In R.C. Atkinson (Ed.), *Studies in Mathematical Psychology*, pp. 358–379. Stanford, CA: Stanford University Press.

Suppes, P. & Zanotti, M. (1976). On the determinism of hidden variable theories with strict correlation and conditional statistical independence of observables. In P. Suppes (Ed.), *Logic and Probability in Quantum Mechanics*, pp. 445–455. Dordrecht, Holland: D. Reidel.

Suppes, P. & Zanotti, M. (1980). A new proof of the impossibility of hidden variables using the principles of exchangeability and identity of conditional distribution. In P. Suppes (Ed.), *Studies in the Foundations of Quantum Mechanics*, pp. 173–191. East Lansing, MI: Philosophy of Science Association.

Suppes, P. & Zanotti M. (1981). When are probabilistic explanations possible? *Synthese*, **48**, 191–199.

Suppes, P., & Zanotti, M. (1982). Necessary and sufficient qualitative axioms for conditional probability. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **60**, 163–169.

Suppes, P., & Zanotti, M. (1984). Causality and symmetry. In S. Diner, G. Lochak, & W. Selleri (Eds.), *The Wave-Particle Dualism*, pp. 331–340. Dordrecht, Holland: D. Reidel Publishing Co.

Suppes, P., & Zinnes, J.L. (1963). Basic measurement theory. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (Vol. 1, pp. 1–76). New York: Wiley.

Tarski, A. (1951). *A Decision Method for Elementary Algebra and Geometry* (2nd ed.). Berkeley, CA: University of California Press.

Tarski, A. (1956). *Logic, Semantics, Metamathematics*. Oxford: Oxford University Press.

Tarski, A. (1959). What is elementary geometry? In L. Henkin, P. Suppes & A. Tarski (Eds.), *The Axiomatic Method*, pp. 16–29. Amsterdam: North-Holland.

Toomer, G.J. (Ed.). (1976). *Diocles on Burning Mirrors.* Berlin: Springer-Verlag.

Toomer, G.J. (Ed.). (1984). *Ptolemy's Almagest.* New York: Springer-Verlag.

Toth, L. F. (1964). *Regular Figures.* New York: Macmillan.

Truesdell, C. (1968). *Essays in the History of Mechanics.* New York: Springer Verlag., (II, 3rd Edition, pp. 102–108, 151–157). London.

Whitehead, A.N. (1919). *An Inquiry Concerning the Principles of Natural Knowledge.* Cambridge, MA: Cambridge University Press.

Whitehead, A.N. (1920). *The Concept of Nature.* Cambridge, MA: Cambridge University Press.

Whittaker, E.T. (1910). *History of the Theories of Aether and Electricity.* London: Longmans, Green.

Wigner, E.P. (1932). On the quantum correction for thermodynamic equilibrium. *Physical Review*, **40**, 749–759.

Wigner, E.P. (1970). On hidden variables and quantum mechanical probabilities. *American Journal of Physics*, **38**, 1005–1009.

Wolfram, S. (1985). Undecidability and intractability in theoretical physics. *Physical Review Letters*, **54,** 735–738.

Wolfram, S. (1986). Random sequence generation by cellular automata. *Advances in Applied Mathematics*, **7,** 123–169.

Yasue, K. (1981). Stochastic calculus of variations. *Journal of Functional Analysis*, **41**, 327–340.

Yule, G.U. (1911). *An Introduction to the Theory of Statistics.* London: C. Griffin & Co.

Zajaczkowska, A. (1956). Experimental test of Luneburg's theory: Horopter and alley experiments. *Journal of the Optical Society of America*, **46**, 514–527.

Zeeman, E.C. (1962). The topology of the brain and visual perception. In M.K. Fort (Ed.), *The Topology of Three Manifolds*, pp. 240–256. Englewood Cliffs, NJ: Prentice-Hall.

This book was typeset with TEX on Turing, CSLI's principal computer at Stanford University, by Laura Tickle, Kaija Lewis, and Emma Pease, in Computer Modern Roman type, designed by Donald Knuth with his digital-font designing program, METAFONT. TEX, which was also created by Knuth, is a trademark of the American Mathematical Society.

AUTHOR INDEX