THE JOURNAL OF PHILOSOPHY

VOLUME CIX, NUMBER 8/9
AUGUST/SEPTEMBER 2012
SPECIAL ISSUE

ASPECTS OF EXPLANATION, THEORY, AND UNCERTAINTY: ESSAYS IN HONOR OF ERNEST NAGEL

Edited by Bernard Berofsky and Isaac Levi

***Published by* The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY
VOLUME CIX, NO. 8/9, AUGUST/SEPTEMBER 2012

FOREWORD

Both editors of this special issue and four of its contributing authors (all but Schervish and Kadane) were students of Ernest Nagel. Nagel's influence on mid-twentieth-century philosophy in America and the English-speaking world was wide and deep. He and his contemporary, Carl G. Hempel, the two most prominent philosophers of science of that epoch, employed the tools of analytic philosophy combined with a deep knowledge of the sciences to illumine classical questions in epistemology and the philosophy of science. Writing within the context of a pervasive positivism, Nagel imbued the discussion with a pragmatic spirit that both refreshed and inspired. (That spirit is conveyed by the beautiful quotation from *Principles of the Theory of Probability* at the beginning of the article by Seidenfeld, Schervish, and Kadane.) Nagel forged an approach to many issues in the philosophy of science that required analytic rigor together with a sophisticated awareness of the latest developments in the sciences. Each original contribution to this volume concentrates upon one of the many issues Nagel addressed—teleology in biology, explanation and theory construction, reduction, and probability. Nagel was a towering figure, but he was so in spite of his small stature and his gentle, caring nature. He was a great teacher, displaying absolute lucidity in his lectures and a gracious attitude of friendliness and support to his students. The editors have chosen to introduce our subject through the excellent tribute to the man and his work by one of the contributors, Patrick Suppes, written in 1994 for the National Academy of Sciences.

BERNARD BEROFSKY

ISAAC LEVI

0022-362X/12/0908-09/469

BIOGRAPHICAL MEMOIR

ERNEST NAGEL: NOVEMBER 16, 1901–SEPTEMBER 20, 1985*

Ernest Nagel was born November 16, 1901, in Nové Mesto, Bohemia (now part of Czechoslovakia) and came to the United States when he was ten years old. He became a naturalized U.S. citizen in 1919, and received his higher education entirely in the United States. In 1923 he received a Bachelor of Science from the College of the City of New York, in 1925 a Master's Degree in philosophy from Columbia University, and in 1931, a Ph.D. in philosophy from Columbia. He spent most of his academic career at Columbia. He was on the faculty there from 1931 to 1970, with the exception of the academic year 1966–67 when he accepted a position at Rockefeller University. From 1967 to 1970 he held the position of university professor at Columbia, and he continued to be active in the intellectual affairs of the university after his retirement, including teaching seminars and courses. Ernest Nagel died in New York City on September 20, 1985.

After his arrival in New York City in 1911, Nagel spent his entire life there, although he and his family regularly spent the summer in Vermont for many years. On January 20, 1935, he married Edith Alexandria Haggstrom, and they had two sons, Alexander Joseph, who is a professor of mathematics at the University of Wisconsin-Madison, and Sidney Robert, who is a professor of physics at the University of Chicago. His wife Edith died in 1988.

During his long and active academic career Nagel received many honors including honorary doctorates from a number of institutions. He was a Guggenheim Fellow in 1934–35 and 1950–51. He was elected to the American Academy of Arts and Sciences in 1954, and to the American Philosophical Society in 1962. In 1977 he was elected to the National Academy of Sciences.

Nagel's many contributions to the philosophy of science are discussed below, but what is most important to emphasize about his more than forty years' association with Columbia University is the central role he played in the intellectual life of Columbia, and more generally, of New York City. To many generations of students he was

* Reprinted with permission from *Biographical Memoirs, Volume 65*, 1994, by the National Academy of Sciences, courtesy of the National Academies Press, Washington, D.C.

0022-362X/12/0908-09/470–78

the outstanding spokesman of what philosophy could offer in terms of analysis of the scientific method, as it is practiced in many different sciences, and in the relation between science and perennial problems of philosophy such as those of causality and determinism. What is important about this influence is that it was not simply students of philosophy, but students of many different disciplines whom he influenced in a way that many of them still remember. He saw his principal role as that of a philosophical critic of ill-conceived notions from whatever quarter they might come. It is this critical spirit of analysis and reflection that he especially communicated to others. He was properly skeptical of philosophical edifices built independent of detailed scientific considerations. But he was equally critical of the writings of scientists who too blithely thought they could straighten out their colleagues on fundamental philosophical questions without proper knowledge of the many issues involved.

His own intellectual mentors were Morris R. Cohen, with whom he wrote the most influential textbook in logic and scientific method published in the period between the mid-1930s and the mid-1950s, and John Dewey, who taught at Columbia for many years and was one of the most important American philosophers in the first half of the twentieth century. Throughout his career Nagel tried to combine the best elements of Cohen's philosophical realism and Dewey's radical instrumentalism.

His closest colleague, personally and philosophically, was probably Sidney Hook, who also taught in New York City for many years, primarily at New York University. Like Dewey and Hook, Nagel also enjoyed the wider arena of intellectual and political life in New York. He wrote extensively for such publications as *Partisan Review* and *The Nation*, as well as for the standard scholarly journals. With these many different interests and engagements he occupied a position, especially in the intellectual life of New York City, that extended far beyond the boundaries of academic philosophy. Within the university Nagel interacted with colleagues in the sciences in a way that was unusual then, and is unusual now, for philosophers. For example, he gave for many years a famous seminar with Paul Lazarfeld on the methodology of the social sciences, which was widely attended by social scientists as well as philosophers at Columbia. His interest in current research in physics continued well into retirement. It is not common practice for philosophers to be elected to the National Academy of Sciences, and there is no special section to which they naturally belong. His election was a tribute to Ernest Nagel's wide-ranging interests and extensive substantive knowledge of many different branches of science. It is fair to say that the range of his scientific interests and knowledge exceeded that of any other philosopher of science of his generation in the United States.

A SURVEY OF NAGEL'S WIDE-RANGING INTERESTS

Nagel wrote about too many different parts of science to survey in detail all that he had to say. What I do want to do, however, is to give a sense of the wide range of his interests and the continual concern for foundational issues discussed from the critical standpoint he thought essential for a philosopher.

Causality, Explanations, and Laws. The general topic of causality, and also the nature of scientific explanations and laws, are topics to which Nagel returned again and again in his career. His most extensive discussion is to be found in his magisterial book, *The Structure of Science,* which has as its subtitle *Problems in the Logic of Scientific Explanation.* Here he devoted a chapter to patterns of scientific explanation with an analysis of four kinds of explanation offered in science: the deductive model, the probabilistic model, the functional or ideological model, and the genetic model, where by "genetic" is meant the study of the historical roots of phenomena. Although he gave a very sympathetic exposition on various occasions of teleological explanations in biology, he favored the classical deductive model as providing the best examples of scientific explanation. However, he also recognized the problems of characterizing what a nontrivial deductive pattern of explanation must be and in various publications went to some length to analyze the various puzzles surrounding this notion. It would probably be generally conceded that the intuitive notion of a nontrivial deductive explanation is still not thoroughly analyzed, and is possibly not a notion that we shall ever put on a completely formal basis. Nagel was also concerned with the logical character of scientific laws. Many of the same puzzles that beset explanations beset characterizing the nontriviality of laws. He was equally concerned to distinguish purely experimental laws from theoretical laws. He had many wise things to say on all of these problems of explanation, laws, and theories without proposing or even believing in some grand general scheme that would satisfactorily account for all the puzzles that have been raised about these concepts. As I have already emphasized, what is important about Nagel's role as a critic of science and philosophy is that he did not focus only on general issues about causality and explanation, but went on to the detailed analysis of these concepts and their use in individual scientific disciplines.

Foundations of Measurement. In his dissertation completed in 1931 and throughout his academic career, Nagel had continuing interest in the theory of measurement. More than any other philosopher of his generation he built on the nineteenth-century work of Helmholtz and Hölder, as well as the earlier-twentieth-century work of the British physicist Norman Campbell. It was characteristic of Nagel's approach that he did not extend the formal results obtained earlier by Hölder and others, but critically examined the conceptual assumptions back of the formal developments.

Foundations of Geometry. Already in his dissertation he exhibited his deep interest in the history of nineteenth-century geometry. He continued this interest in a number of publications; one of his most well-known pieces of work is a detailed examination of the development of the conception of systems of geometry as abstract mathematical structures in the nineteenth century. The central role that geometry played in the development of the abstract form of modern mathematics has often not been appreciated sufficiently in discussions of the foundations of mathematics by mathematicians and philosophers. The development of projective geometry by Monge, Poncelet, Gergonne, Von Staudt, and others, as well as the abstract theory of Grassman's *Ausdehnungslehre* and related work, formed the background for the rapid development of the modern axiomatic view of geometry developed by Pasch, Hilbert, and Klein in the last decades of the nineteenth century. Nagel's long essay, published in 1939, was one of the first historical analyses to recognize the great importance of the break that was made by the introduction of projective geometry for later views on the foundations of mathematics. What was essential was the new understanding that pure geometry is neither the science of quantity nor the science of extension in the sense so thoroughly developed by Euclid.

Years later, Nagel took up again his interest in geometry, in the chapter devoted to space and geometry and in another chapter to geometry and physics, in *The Structure of Science*. He analyzed with care the foundational discussions of the differences between pure and applied geometry and the nature of conventions in geometry, with particular reference to the much earlier discussions by Poincaré and Einstein. Nagel presented persuasive arguments why Poincaré was wrong in his judgment that Euclidean geometry would never be abandoned.

Foundations of Physics. As already indicated, Nagel devoted a substantial part of his critical energy to the fundamental philosophical issues raised by the development of relativity theory and quantum mechanics during the period spanned by his academic career. His concern to give a detailed philosophical critique of the relation between geometry and physics was just mentioned. The issues raised by quantum mechanics were of equal importance to him. In various publications he was concerned to distinguish the sense in which quantum mechanics preserves causality as reflected in the deterministic solutions of the Schroedinger equation for given initial conditions, and at the same time to analyze the many different senses in which quantum phenomena could be said to be indeterministic. He was very much aware of the fact that there is no single sense of indeterminism that is agreed upon as the central one, and also that different senses of indeterminism depend upon different senses of the concept of probability. Here is a characteristic passage from Nagel's writings on the matter: "In the voluminous literature on the 'indeterminism' of microphysics, one point stands out clearly: whatever the issue may be, it is generated by the theoretical interpretations that are placed on the acknowledged data rather than by any disagreement as to what those data are."

Another classic paper of Nagel's is concerned with the detailed analysis of the reduction of theories, with special emphasis on the reduction of thermodynamics to statistical mechanics. This is a subject that has received much attention from applied mathematicians and theoretical physicists in the last half century. Nagel does not add to the technical results on the complex problem of giving clear mathematical results concerning under what conditions a representation theorem can be proved, but he does provide the most extensive conceptual analysis to be found over a long period in the literature of the philosophy of science on this important case of reduction. More generally, his analysis of the reduction of theories in a chapter of *The Structure of Science* is a classical presentation of philosophical views on reductionism.

Foundations of Probability. Throughout the twentieth century there was extended conceptual controversy over the nature of probability. The terrain of the conflict has not been restricted to any one domain of science, although physics has been central to much of the discussion, but equally important has been the Bayesian view that the most important sense of probability is the subjective one of degree of belief, advocated most persuasively by Bruno de Finetti and L. J. Savage. Most of Nagel's writings on the foundations of probability appeared before Savage's 1954 book, *The Foundations of Statistics*. Although Nagel vigorously defended the frequency interpretation of probability, he was careful to survey the various logical problems that have been raised about the frequency interpretation, including well-known objections to Von Mises's concept of a collective. He was also among the first in the philosophical literature to call attention to the important method of arbitrary functions in probability theory, developed to provide an account of physical mechanisms in coin flipping and other such physical devices for producing symmetric probability distributions. He acknowledged especially the important work of Poincaré, carried on later by G. D. Birkhoff, E. Hopf, and others, providing a detailed account of the ordinary physical mechanisms by which symmetric probabilities are produced in games of chance such as roulette, craps, and so on.

Theories of Induction. Much more of Nagel's intellectual energy was devoted to critical analyses of theories of induction put forth, especially by the philosophers Hans Reichenbach and Rudolf Carnap, who made proposals sufficiently detailed to also attract the attention of statisticians interested in the foundations of statistical inference.

Although agreeing with Reichenbach that the relative frequency interpretation of probability is the fundamental one, Nagel on numerous occasions criticized Reichenbach's wholesale attempts to extend the relative frequency theory to give an account of the quantitative degree of confirmation of a scientific theory. Nagel rightly believed that Reichenbach's efforts in this direction were too crude and general to provide a serviceable methodology for evaluating the probability of a theory. Nagel's characteristic skepticism of philosophers who propose simple and general theories for complex matters comes through again and again in his criticisms of Reichenbach's ideas. It is fair to say that Reichenbach's analysis no longer has serious currency. Nagel's published criticisms were one of the most effective lines of attack against Reichenbach's far too sweeping proposals.

Nagel criticized in a similar fashion Reichenbach's unorthodox and equally sweeping proposals for the interpretation of quantum mechanics. For example, Reichenbach proposed a three-valued logic of true, false, and indeterminate, but did not provide anything like the proper intuitive and technical development of this logic. Nagel's criticisms were characteristically sharp and pointed.

With equal claim to generality but with a completely different interpretation of probability, namely what is usually termed a logical theory of probability, Rudolf Carnap proposed a general approach to the theory of confirmation of scientific theories. Nagel managed to find as many intuitive difficulties with Carnap's theory as with Reichenbach's. What is important to record here is not the technical criticisms of Carnap or Reichenbach, but rather the general perspective from which he conducted these critical investigations. He clearly felt that the effort to have a general methodology for quantitative confirmation of scientific theories, taken as wholes, was an unworkable and unfeasible idea. Drawing upon his own wide scientific knowledge he offered numerous counter-examples to Carnap's ideas. Nagel was equally critical of the fact that Carnap based his theory of induction on assuming that we were able to characterize a set of independent and complete primitive predicates for describing experience. Nagel puts his criticism this way: "it is difficult to avoid the conclusion that the assumption that we have, or some day shall have, a complete set of primitive predicates is thoroughly unrealistic, and that in consequence an inductive logic based on that assumption is a form of science fiction."

Scientific Explanation in Biology. Over a period of many years, Nagel published a number of articles on the character of scientific explanations in biology. He included in *The Structure of Science* a chapter on mechanistic explanation and organismic biology, and in the John Dewey lectures, given at Columbia University in 1977, he gave perhaps his most thorough analysis of the concept of teleology in biology. Nagel's Dewey lectures provided a reformulation and reexamination of his earlier writings on teleological explanation. The written version of the lectures is divided into two parts. In the first part Nagel examined three alternative accounts of the notion of goal and goal-directed processes. The first is the intentional account, which is modeled on purposive human behavior, and, rightly enough, Nagel finds difficulties with this view in talk about goal-directed processes in lower organisms such as protozoa and plants. The second account is the computer-program view of such processes; genetic coding is a striking and appealing example, but Nagel points out that the concept of goal-directedness is one that we attribute to behavior without having the possibility of examining any proposed internal computer program that controls it.

The third account of goal-directed behavior Nagel refers to as the "system-property" view of goal-directed processes. An example that illustrates this view is the collection of mechanisms that act homeostatically to maintain the water content of the blood at about ninety percent. Nagel imposes the reasonable requirements that the process be plastic, that it be persistent, and that the relevant variables controlling it be for the normal range of their values independent. It should be obvious that there is no inconsistency between the computer-program view and the system view, but it is the system view that he uses for the definition of goal-directed behavior, for the reason already indicated. Nagel also deals with several objections to the system view which I shall not examine here. The important point is that once the system-property view is accepted, then a general analysis of the concept of being goal-directed can be given without using specifically biological notions or other expressions that have a teleological connotation. By giving an analysis of goal-directed processes in this fashion, Nagel wanted to make the important point that explanations of goal-directed processes in biology are, in principle, similar in structure to explanations of nonbiological processes in the physical sciences.

The second part of the essay is devoted to functional explanations in biology, the second main type of teleological explanation. Nagel says that a typical example of a functional explanation is the assertion, "fish have gills in order to obtain oxygen." The basic form of functional explanations for Nagel is this: "During a given period t and in environment E, the function of item i in system S is to enable the system to do F." An example would be green plants being provided during a period of time, water, carbon dioxide, and sunlight, with the function of chlorophyll then being to enable the plants to perform photosynthesis. As Nagel notes, such functional explanations are not causal, in contrast to explanations of goal ascriptions. In the process of setting forth his own views, Nagel examines Carl Hempel's well-known critique of functional explanations and defends a proper formulation of their use in biology.

Methodology of the Social Sciences. Nagel's general thesis about the social sciences is that they are subject to the same general canons of scientific method applicable in the natural sciences. He was particularly concerned to argue on numerous occasions that subjective explanations of human behavior either individually or in groups—an approach that has a long history of proponents—does not satisfy the usual standards for scientific inquiry and can be avoided. He dealt in the same way with the claims that investigations in the social sciences are subject to a peculiar form of value-oriented bias. In various publications Nagel was also concerned to offer a detailed analysis of the nature of statistical explanations in the social sciences, especially emphasizing their importance for causal analysis. Finally, I would not want to omit the fact that he devoted the last chapter of *The Stucture of Science* to problems in the logic of historical inquiry. He provided in this final chapter a particularly careful and detailed analysis of three important problems: the problem presented by the selective character of historical inquiry for the achievement of historical objectivity; the scientific justification for assigning relative importance to causal factors, as for example, the relative weight of economic as opposed to political factors as causes of the American Civil War; and finally, the possibility of using effectively in history contrary-to-fact judgments about the past, in order to evaluate the nature of various historical events.

FINAL NOTE

I have surveyed in a necessarily superficial way Ernest Nagel's many philosophical and scientific interests. What is equally important is to emphasize the unity of his vision of the nature of scientific inquiry and the critical role that philosophy of science can have in rooting out mistaken conceptions and ill-thought-out claims of significance. Because of the emphasis he placed on criticism, it is not possible in any simple way to summarize the unity of Ernest Nagel's intellectual vision. However, an easily identified style and manner of thought come through in his writings in any of the areas I have surveyed. The same patient critical tone permeated his seminars as well as his written work. As legions of students will attest, a seminar or course with Ernest Nagel was a memorable experience, perhaps above all because his persistent criticisms were tempered by a rare gentleness of personality and spirit.

PATRICK SUPPES

Stanford University

THE EXPLANATION OF LAWS: SOME UNFINISHED BUSINESS

> Even if we were sure that all possible laws had been found and that all the external world of nature had been completely ordered, there would still remain much to be done. We should want to *explain* the laws.[1]

Norman Campbell rightly set the task. It is the business of science not only to discover laws, but to explain them. And he added his voice to a philosophical tradition going back to Aristotle, of taking on the task of explaining what laws are, and explaining as well what explanations of laws are. Ever since the publication of the seminal paper of Hempel and Oppenheim on scientific explanation,[2] philosophers have been inspired to do better on the subject. But it became painfully clear, from the counter-example Hempel and Oppenheim offered in their paper, that their account of scientific explanation could not cover the explanation of laws. Although this is the business of philosophers, it is still unfinished business.

I. THE CAMPBELLIAN BACKGROUND

Ernest Nagel and Richard Bevan Braithwaite were well aware of Campbell's views on the structure of theories, and referred to them when they addressed the issue of the explanation of laws directly. Both had, as we shall see, very different accounts of laws, and their explanations. Braithwaite (1953) developed a view that can be traced back to J. S. Mill and F. P. Ramsey (a view which Ramsey later rejected) that was a very different variation of the Mill-Ramsey view that David Lewis developed some two decades later (1973). Nagel developed a novel view that is an interesting combination of the views of Norman Campbell and David Hilbert. Sad to say, however, neither of these remarkable accounts got the timely critical attention that they deserved. Our present task, our unfinished business is to revisit them, and perhaps generate new interest in them.

[1] Norman Campbell, *What Is Science?* (London, UK: Methuen and Co. Ltd., 1921), p. 77.

[2] Carl G. Hempel and Paul Oppenheim, "Studies in the Logic of Explanation," *Philosophy of Science*, xv, 2 (April 1948): 135–75.

0022-362X/12/0908-09/479–502

II. DEDUCTIVE SYSTEMS AND THEIR LAWS

In a very laudatory review of Richard Braithwaite's *Scientific Explanation*,[3] Ernest Nagel raised an objection to Braithwaite's account of scientific laws which is worth quoting in full:

> Braithwaite's explicit formulations of the structure of explanatory systems…do not always make entirely clear whether he thinks that the deduction of an empirical generalization from some established higher-level hypothesis is sufficient to constitute an explanation for the generalization….In any event, it is doubtful whether most physicists would accept as an "explanation" of, say, Galileo's law for freely-falling bodies ($d = gt^2/2$) the derivation of this law from the logically equivalent hypothesis that $t = \sqrt{(2d/g)}$. Deducibility from higher-level hypotheses is at best only a necessary condition for explanation, not a sufficient one. The explanatory hypotheses in most of the actual scientific systems are required to meet other conditions as well: possessing a greater "generality" (not to be identified with greater deductive power) than the lower-level hypotheses[4]….It would improve his exposition, nevertheless, had he discussed them systematically and so guarded himself against the possible misconception that the sole task in the quest for scientific explanations is the relatively easy one of constructing a deductive system in which empirical generalizations are theorems.[5]

The example of two logically equivalent versions of Galileo's law of freely falling bodies is beside the point. In the kind of scientific deductive system that Braithwaite considered (something we shall try to explain below), it would not be acceptable to have an hypothesis be a higher-order hypothesis, and its logical equivalent be a lower-order hypothesis in the same deductive system. Nevertheless Ernest's point is a good one. It does seem that typical examples of an explanation of laws involve a least a premise of greater generality.

The second admonition by Ernest is that there is the impression—to be guarded against—that explanation of laws amounts to showing them to be theorems in a deductive system. Ernest thought that providing explanations, on that count, would be a relatively easy task. However Braithwaite's proposal for distinguishing the *laws* of a well-established deductive system is simply this:

> The condition for an established hypothesis *h* being *lawlike* (i.e., if true, a natural law) will then be that the hypothesis either occurs in an established

[3] Richard Braithwaite, *Scientific Explanation: A Study of the Function of Theory, Probability, and Law in Science* (Cambridge, UK: University Press, 1953). All page references for Braithwaite in the main text are to this book.

[4] We omit the other features Nagel mentioned since they concern analogy and evidential requirements that are not directly relevant to the issues under discussion.

[5] Ernest Nagel, "A Budget of Problems in the Philosophy of Science," *Philosophical Review*, LXVI, 2 (April 1957): 205–25, at p. 206.

> scientific deductive system as a higher-level hypothesis containing theoretical concepts or that it occurs in an established scientific deductive system as a deduction from higher-level hypotheses which are supported by empirical evidence which is not direct evidence for *h* itself. (301–02)

It is not so evident that such *established* deductive systems are relatively easy to come by. At any rate, in this passage and others, it is clear that Braithwaite proposed conditions for something to be a law. Braithwaite's conditions for something to be an explanation of a law are another matter entirely. Ernest however took these conditions to be a very different proposal: one for something to be an explanation of a law, rather than a proposal for what counts as a law.

As I understand Braithwaite's project, the idea is to single out those generalities of an established deductive system which are its laws. His proposal could be stated this way:

> Let *S* be some generalization of some established deductive system *D*. Then *S* is a law of *D* if and only if there is some generalization *S** of *D* which implies *S* (but not conversely).[6]

The place of a generalization in a deductive system is the key ingredient in sorting out the laws from the other generalizations of the system. Braithwaite's reliance on the place that a generalization has in a deductive system is an unusual requirement. The usual accounts of laws usually make no reference to position in a structured body of statements. For example, there is the requirement that laws are those contingent generalizations that are not accidental. There is also the claim that laws are those generalizations that support their corresponding counterfactuals. Then too, there is the account according to which laws are those contingent generalizations that are physically necessary. None of the notions of "accidental generalization," "counterfactual conditional," or of "physical necessity" appeal to some structured body of statements.[7]

[6] We have added the parenthetical condition to avoid the consequence that every generalization of an established deductive system would be a law. It still does not blunt the possible case where a deductively strongest generalization of a deductive system might not be a law of that system. This is a consequence which Braithwaite called to attention, and for which he developed an answer. It should also be noted that although such a highest generalization might not count as a law of a system *D*, it might very well be a law of an established deductive system that was an extension of *D*. His additional requirement that a highest-order hypothesis in a deductive system that has an occurrence of a theoretical term is a law seems very *ad hoc* to me. Nevertheless there are several important plausible cases when this is so—in systems in which the highest-order hypotheses are the three laws of Newtonian mechanics, Newton's theory of gravitation, Schrödinger's formulation of quantum mechanics, and the basic (three or four) laws of classical thermodynamics.

[7] I am not endorsing any of these accounts. In fact there are fairly convincing examples of accidental generalizations that also support their corresponding counterfactual

The explanation of laws is another closely related matter which we will examine in more detail later in this essay. We shall simply for the present call attention to two seminal paragraphs that capture Braithwaite's views on explanations of laws.

> (1) To explain a law, as we have seen, is to incorporate it in an established deductive system in which it is deducible from higher-level laws. To explain these higher-level laws is similarly to incorporate them and the deductive system in which they serve as premises, in an established system which is more comprehensive and in which these laws appear as conclusions. To explain the still-higher-level laws will require their deduction from laws at a still higher level in a still more comprehensive system. (347)
>
> (2) Any incorporation of a fact—be it a particular instance of a law or the law itself—into a deductive system in which it appears as a conclusion from other known laws is, by virtue of that incorporation, an explanation of that fact or law.…what matters is that we know more than we did before of the connectedness of the fact or law with more fundamental laws covering a wider range. (349)

It is however certainly true, as Nagel noted, that an explicit account of one law's being more general or more fundamental or having wider range is patently missing from this account of the explanation of laws. This we shall try to rectify by a slight addition.

Braithwaite's account of the explanation of laws in (1) and (2) relies upon the use of established deductive systems and is very different from his account of laws. We turn then to his account of laws, and subsequently to his account of their explanation, after some remarks about his notion of a deductive system.

III. DEDUCTIVE SYSTEMS AND THEORIES

It is not at all obvious what Braithwaite meant by "deductive system." Although influenced by Campbell's account of theories, which we shall describe more fully below, deductive systems are not theories in Campbell's sense. Here is Braithwaite's simplified example—a "Gallilean" deductive system, based roughly on Galileo's law of falling bodies (13). Informally the statements in the system are arranged in levels with those on the higher levels

---

conditionals. Even worse, we think that it can be shown that if a law implies its corresponding counterfactual conditional, then it is equivalent to that counterfactual. In that case there are examples of laws which scientific practice regards as logically equivalent, while their corresponding counterfactuals are not equivalent.

implying (but not being implied by) those on the lower. The idea was that

> The hypotheses in this deductive system are empirical general propositions with diminishing generality. (13)

The system he had in mind is illustrated by the following statements in descending order of generality:

(I) Every body near the Earth freely falling towards the Earth falls with an acceleration of 32 feet per second per second.
(II) Every body starting from rest and freely falling towards the Earth falls $16t^2$ feet in t seconds, whatever number t may be.
(IIIa) Every body starting from rest and freely falling for 1 second towards the Earth falls a distance of 16 feet. (And so on for other values of t).

It is obvious from this example that a deductive system cannot be identified with a set of axioms from which all the other statements in it follow (the elements in (IIIa) are not axioms of the system), nor is it the set of all logical consequences of a set of axioms (logical truths are not in it). So deductive systems are neither a theory in the sense of a set of axioms like the axioms for Euclidean geometry, nor are they a theory in the sense of Tarski (the set of logical consequences of some set of sentences).

Probably one way to represent Braithwaite's deductive systems would be as a lattice (not Boolean, since that would restrict systems to those having exactly one higher-order hypothesis). That would allow one to place hypotheses of various generality at various nodes with arrows indicating a type of implication.

A more formal description of these deductive systems is not a serious problem. Here is one suggestion: what is needed is a notion of a structure that contains more than the listed axioms of a theory in the first sense of "theory" we discussed, and less than the full Tarskian theory. Define a Braithwaitean deductive system D^* as any set of statements which satisfies the following conditions, where S and S^* are any contingent hypotheses (generalizations):

(1) D^* contains all the axioms of some theory T;[8]
(2) If S logically implies S^* ($S \Rightarrow S^*$), then if S is in D^*, so too is S^*;
(3) If S is in D^*, and I is an instance of S, then I is in D^*;

[8] We believe that here Braithwaite would have used something close to Campbell's notions of the hypothesis and the dictionary of a theory. Those notions will be discussed more fully below.

where (2) is a kind of closure condition restricted to contingent generalizations. It is clear however that although the structure of the system involves deductions, it also involves levels of generality, and an account of "is more general than" is missing. We shall see below that it is Nagel, and I believe only Nagel, who has offered an explicit account—actually two—of "is more general than," one more general than the other (no pun intended).

It is clear that both Braithwaite and Nagel thought that something like greater generality has to figure in the explanation of laws. The requirement as Nagel expressed it is that for the explanation of laws, at least one of the premises has to be more general than what is explained. Of course the requirement does not seem to be needed for explanations of singular facts. Neither does it figure in those explanations of events that consist in locating them in a causal network. The fact that some notion of generality figures in explanations of laws but not in other kinds of explanations is worth worrying about—only not here.[9]

As we noted above, the account of laws that situates them in deductive systems is unusual. This use of deductive systems has an interesting history, and before we turn to Braithwaite's account of the explanation of laws, it is important to consider the problem that Braithwaite attempted to solve with this account.

IV. THE MILL-RAMSEY-BRAITHWAITE ACCOUNT OF LAWS (MRB)

The story of the origin of the Mill-Ramsey-Lewis (MRL) account of laws begins with some remarks of the Cambridge philosopher W. E. Johnson about two kinds of conditionals—universals of fact, and universals of law. That prompted a sharp reply by another Cambridge philosopher, F. P. Ramsey. Ramsey then entertained an account that essentially had been proposed earlier by J. S. Mill—the Mill-Ramsey account (MR). Soon thereafter, Ramsey (and, I am fairly sure, Braithwaite) rejected (MR). Braithwaite then developed a better version of it. Some two decades after that, David Lewis developed a different nonepistemic version (MRB) of the discarded Mill-Ramsey account. It is the remarkable but neglected Braithwaite account that we will now describe.

Ramsey's sharp reaction to Johnson's distinction between two kinds of conditionals was that it was impossible. Johnson thought that universals of fact were universal quantifications over all things that in fact satisfied a conditional, whereas with universals of law, the

[9] I suspect that part of the reason is that "is more general than" was sometimes used interchangeably with "is more comprehensive than" and with "is more fundamental than."

quantifier ranged over all possible things that satisfied a conditional. Ramsey (unpublished note) remarked that Johnson got the quantifiers wrong and failed to realize that "everything" means everything. Ramsey then considered an account of laws that J. S. Mill had advocated:

> …What are laws of nature? May be stated thus: What are the fewest and simplest assumptions, which being granted, the whole existing order of nature would result? Another mode of stating the question would be this: What are the fewest general propositions from which all the uniformities which exist in the universe might be deductively inferred?[10]

Ramsey's compact expression of that view was simply that

> Laws are consequences of those propositions which we should take as axioms if we knew everything and organized it as simply as possible in a deductive system.[11]

However Ramsey (and I am fairly sure Braithwaite) rejected it on the epistemic grounds that it is impossible to know everything and to organize it in a deductive system.

The Millean proposal then was developed in at least three different ways. Ramsey went on to consider laws as variable hypotheticals—a device which is not even propositional. I shall say no more about this possibility. Another better-known response is due to David Lewis, who expunged the epistemological element from Ramsey's formulation, replaced the deductive system of the totality of everything known by another deductive system, and accounted for laws this way:

> …a contingent generalization is a law of nature if and only if it appears as a theorem (or axiom) in each of the deductive systems that achieves a best combination of simplicity and strength.[12]

With this variation on the Mill-Ramsey theme we have one of the most durable and plausible Humean accounts of law in contemporary philosophy of science—(MRL).

With the Mill-Ramsey version (MR), we had a deductive system, some of whose propositions were known to be in it, but certainly not all. With Lewis's variation, we have an appeal to a deductive system for which it is not a sure thing that one's favorite generalizations will

[10] John Stewart Mill, *A System of Logic, Ratiocinative and Inductive* (New York: Harper and Bros., 1858), p. 230.

[11] F. P. Ramsey, *Philosophical Papers*, ed. D. H. Mellor (New York: Cambridge, 1990), p. 150.

[12] David K. Lewis, *Counterfactuals* (Cambridge: Harvard, 1973), p. 73. A more nuanced version can be found in his "Humean Supervenience Debugged," *Mind*, CIII, 412 (October 1994): 473–90.

be included in it. That is, of all the true deductive systems that are possible, the laws will be members of the one which is the best combination of simplicity and informativeness. Even if it were settled how simplicity and informativeness were to be construed, it is less than satisfying if the answer to the question "Is *P* a law?" for any true generalization *P*, is: "I don't know," "I'm not sure," or "It's anyone's best guess." The epistemic reservations raised by Ramsey remain in force.

Braithwaite's proposal (MRB) differs significantly from Lewis's later variation. Lewis expunged the epistemic drawback of the Mill-Ramsey proposal. Braithwaite's idea was to meet, in one fell swoop, the two objections of Ramsey to the Mill-Ramsey account (that it is impossible to know everything and to organize it in a deductive system) by using epistemically more modest candidates. Instead of one axiomatization of all knowledge, he proposed to use instead *established* deductive systems organized around specific Campbellian theories.[13]

V. DEDUCTIVE SYSTEMS AND THE EXPLANATION OF LAWS

If we turn to the first of the two paragraphs in which Braithwaite explicitly described the condition for an explanation of a law, we find an intricate use of deductive systems and their proper extensions. That paragraph, (1), can be described as a compact condition on explanations this way:

(E) Let *A* be a law in the deductive system *D*. *A* is explained by *L* if and only if *L* is a law in some deductive system D^* that is a proper extension of *D*, and *L* implies, but is not equivalent to, *A*.

Basically the idea is that you cannot explain a law in a deductive system by restricting yourself to the members of that system. You have to use a law (or laws) in a proper extension from which it follows. One of the best examples of explanation (one of several) for him was the explanation of an approximate formulation of Newton's law of gravitation by Einstein's general theory of relativity.

In the second paragraph, (2), he speaks of the explanation of a law as the incorporation of it into a deductive system. We take him to mean that it is a case of the incorporation (we prefer to say "the embedding") of one entire deductive system within another, *where the explaining law is not a member of the embedded system.* Braithwaite thought that the larger system had greater predictive power. To that extent he probably thought of the larger system as more

[13] One interesting difference between (MR) and (MRL) on the one hand, and (MRB) on the other, is that all the statements of the first two kinds of deductive systems are true, but the statements of (MRB) only need to be established.

comprehensive. However that systemic consideration does not lead, as far as I can see, to any notion of one law being more general than another. Nagel's condition that any explanation of a law requires the use of some more general law is not vindicated on the Braithwaite account. Nevertheless that account has some significant features that we can only mention at present:

(1) It does not fall victim to the Hempel-Oppenheim counter-example to their own Deductive-Nomological Model, when applied to laws. That model would require that the conjunction of Kepler's laws of planetary motion and Boyle's law of gases explains Kepler's laws. It can be shown that that absurd example is not permitted according to Braithwaite's proposal.

With Braithwaite-style deductive systems in place, there are some very nice consequences which should be noted:

(2) (LL) If *L* is a law of a deductive system *D*, and *A* is a contingent generalization that is logically implied by *L*, then *A* is also a law of *D*.

Thus, contingent generalizations which follow from laws of *D* are also laws of *D*.[14] So "It is a law of a deductive system that…" distributes over implication. There are also some nice formal properties of explanation:

(3) (EL) If *A* is a contingent generalization that is explained in some deductive system *D* then *A* is also a law of *D*.

So, if you think that laws are important, then explanations are one way of assuring that explained generalizations will also be laws. Lastly, there is a consequence, similar to (LL), for explanations of laws.

(4) (EE) If *A* is a contingent generalization of the deductive system *D* that is explained in the deductive system D^*, and A^* is a contingent generalization implied by *A*, then A^* is also explained in D^*.

The Braithwaite variation on the Mill-Ramsey proposal has not been recognized for what it is—a Humean account of laws and their explanations, that exploits the relation that laws have to established deductive systems that contain them, and connects the explanation of laws to the embedding of established deductive systems within more comprehensive ones. The connection with scientific practice is evident, and the formal features it supports we believe heightens its interest. We believe that it is a step in the right direction, and

[14] The more general result also holds: If L^* is a generalization that follows from several laws of *D*, then it too is a law of *D*.

find it disappointing that it has not been given the serious scrutiny that it evidently deserves.

VI. ERNEST NAGEL'S THEORY; CAMPBELLIAN ORIGINS

Ernest's account of the explanation of laws is deeply indebted to Campbell, especially for the novel way in which the Campbellian representation of theories is deployed to yield a radical account of the explanation of laws. It too has been neglected, perhaps because it was mistakenly viewed as something already familiar from the published works of Hempel and others—the so-called "received view." Nothing could be further off the mark.

Nagel thought that all explanations of laws were deductive (33).[15] Drawing upon a detailed example of an explanation of the law that ice floats on water, here is his compact description of at least three conditions that would hold for any explanation of laws (including statistical ones):

> …all the premises are universal statements; there is more than one premise, each of which is essential in the derivation of the explicandum; and the premises taken singly or conjointly, do not follow logically from the explicandum. (34)

Two caveats: We will not be concerned with Nagel's discussion of the explanation of statistical laws.[16] We will also separate the cases when the explanation of a law involves only "empirical" laws as premises (they contain no occurrences of theoretical terms), from those in which at least one "theoretical" assumption is used.[17] Nagel assumed that it made no difference—though, as we shall explain below, we think it does.

With some important differences, Nagel adopted Campbell's canonical representation of theories as given by two mutually exclusive sets of statements. The first, the *hypothesis* of the theory, contains those statements of the theory whose only occurrences of nonlogical terms are theoretical. The second, the *dictionary* of the theory, contains those statements that have occurrences of both theoretical and

[15] All subsequent page references in the main text are to Nagel, *The Structure of Science: Problems in the Logic of Scientific Explanation* (New York: Harcourt, Brace and World, 1961).

[16] His conclusion, after a review of typical examples, is that explanations of statistical laws are deductive, at least one premise is statistical, and at least one premise must have a greater degree of statistical dependence than that of the law to be explained. *Ibid.*, p. 520.

[17] Nagel was certainly aware of the various criticisms of such a distinction between theoretical and observational terms, but he has a very vigorous account of that distinction, admitting its vagueness, that cannot be discounted.

observational terms. It is important to note that for Campbell, the hypothesis and the dictionary consist of statements which have a truth-value. Nagel adopted this canonical description of theories with one crucial difference. Nagel associated three components with each theory:

(1) an abstract calculus that is the logical skeleton of the explanatory system, and that "implicitly defines" the basic notions of the system;
(2) a set of rules that in effect assign an empirical content to the abstract calculus by relating it to the concrete materials of observation and experiment; and
(3) an interpretation or model for the abstract calculus, which supplies some flesh for the skeletal structure in terms of more or less familiar conceptual or visualizable materials. (90)

Clearly, Nagel intended not only to reflect but to improve upon Campbell's canonical representation of how theories explain laws.

For Campbell, the representation of the required deduction would look roughly like this: a finite number of hypotheses relating only some of the theoretical terms to other theoretical terms, a finite number of dictionary entries which Nagel called "coordinating definitions" (Hans Reichenbach's terminology) relating some of the theoretical terms to some observational ones, and finally the logical conclusion, some law *L*.

VII. EXPLANATION AND THEORIES WITHOUT TRUTH-VALUES

There is no point in trying to indicate the forms that the sentences in each group might have in order to insure that the deduction is correct. There are too many theories of various forms to do that. Nagel's representation of the explanation of a law would have these features: the premises would include some members of the abstract calculus of the theory, together with some statements that correspond to what Campbell called the dictionary of the theory, followed logically by the conclusion—the law to be explained. Nagel calls the dictionary items "rules," relating theoretical to observational terms, but I do not think it matters much for the following argument whether we use the statement or the rule version for the dictionary entries. So, roughly, Nagel's representation of the explanatory argument of a law should look something like this simple example (*modulo* the point that most everyone agrees to, that in such explanations there would usually be two hypotheses):

(i)	$T(\tau_1, \tau_2)$	(a theoretical statement)
(ii)	$D(\tau_1, o_1)$	(a dictionary entry)
(iii)	L.	(the law to be explained)

By focusing on Nagel's construal of (i) and (ii), it will become apparent that Nagel had in fact proposed a bold variation on Campbell, that resonates with a tradition that goes back at least to the geometer Moritz Pasch, and more clearly to a notion of axiomatization that David Hilbert advocated for mathematics and the physical sciences.

On Campbell's account, (i) is an hypothesis relating two theoretical terms, and hypotheses are statements that have a truth-value. In contrast, Nagel thought that (i) was not a statement at all, and was neither true nor false. Campbell regarded the dictionary entry (ii) of the theory as a statement that had a truth-value. In Nagel's version, the coordinating definitions are now described as rules. For that reason, they too are neither true nor false.[18]

Ernest's reason for taking Campbell's hypotheses as lacking truth-value is simple enough, though Ernest never made the argument explicit. I believe it rested on two assumptions. The first is that he thought that an "abstract calculus" (his version of Campbell's hypotheses of a theory) was an axiomatization of the theory, and second, he also thought that axiomatizations of theories were to be understood in a way that Hilbert made famous with his special concept of axiomatization in mathematics and the sciences. That, view, as we shall see, involved understanding that the axioms of theories failed to be statements, and thus were neither true nor false.

VIII. EXPLANATION: AXIOMATIZATION AND THE HILBERT CONNECTION

Although Ernest did not argue for these assumptions, the reasons he held them seem to me to be evident. If we look to the examples that he provided for those theories he called "abstract calculi," three are singled out: Euclidean geometry, the kinetic theory of gases, and probability theory as axiomatized by A. Kolmogorov. Concerning Euclidean geometry, Nagel said that

> The postulates of the system are frequently stated with the expressions 'point,' 'line,' 'plane,' 'lies between,' 'congruent with,' and several others as the basic terms.…Indeed, in order to prevent the familiar although vague meanings of those expressions from compromising the rigor of proofs in the system, the postulates of demonstrative geometry are often formulated by using what are in effect predicate variables like '*P*' and '*L*,' instead of the more suggestive but also more distracting descriptive predicates 'point' and 'line.' (91–92)

[18] Consequently, (ii) should be listed as a rule, rather than as a premise. Whatever status it has, it is supposed to guarantee a deductive transition to the conclusion, since, for Nagel, Campbell, and Braithwaite, all explanations are deductive.

The example of probability theory is especially interesting, because it makes evident the connection of Ernest's view of these axiomatizations with Hilbert's axiomatization of Euclidean geometry. Kolmogorov said explicitly that in providing axioms for probability, he was trying to do the same for that theory that Hilbert did for Euclidean geometry. Nagel was acutely aware of the development of this Hilbertian view of the axiomatization of Euclidean geometry and its earlier anticipation by Pasch. In his penetrating early study of the development of geometry he noted:

> Indeed [Pasch declares], if geometry is to be really deductive, the deduction must everywhere be independent of the *meaning* of geometrical concepts, just as it must be independent of the diagrams; only the *relations* specified in the propositions and definitions employed may legitimately be taken into account. During the deduction it is useful and legitimate, but in *no* way necessary, to think of the meanings of the terms; in fact, if it is necessary to do so, the inadequacy of the proof is made manifest. If, however, a theorem is rigorously derived from a set of propositions—the *basic* set—the deduction has a value which goes beyond its original purpose. For if, on replacing the geometric terms in the basic set of propositions by certain other terms true propositions are obtained, then the corresponding replacements may be made in the theorem; in this way we obtain new theorems as consequences of the altered basic propositions without having to repeat the proof.[19]

This insight is incorporated into the very heart of Nagel's account of explanation of laws by theories. The upshot of these considerations was to replace terms like "point" and "line" in a theory of Euclidean geometry by predicate variables—say "P" and "L." This rewrite has the result that (i), which looked like a statement having a truth-value, has now been replaced by a statement-form. The idea is to do the same for the basic predicates and relations of the kinetic theory of gases, and probability theory. Something quite radical results when this Hilbertian insight is introduced into the original Campbellian representation of theories.

Consider what happens to the explanation of a law in which the deduction is given by the proof-sequence (i) to (iii), once we replace

[19] Reprinted in Nagel, *Teleology Revisited and Other Essays in the Philosophy and History of Science* (New York: Columbia, 1979), pp. 237–38. Originally published in Nagel, "The Formation of Modern Conceptions of Formal Logic in the Development of Geometry," *Osiris*, VII (1939): 142–223. The particular passage (translated by Nagel) is also reprinted in Patrick Suppes, *Representation and Invariance of Scientific Structures* (Stanford: CSLI, 2002), p. 46, where it is used to support a set-theoretical account of axiomatic theories.

(i) and (ii) by the statement-forms (i)′ and (ii)′. The explanatory proof then is given by

(i)′ $T(X_1, X_2)$
(ii)′ $D(X_1, o_1)$
(iii) L,

where the theoretical terms have been replaced by predicate variables. The results are radical, and initially implausible. The first thing to note is that the deduction fails to be an explanation of L. The reason is that explanations are factive—that is, the sentences that are used in explanations (whether as premises, or conclusions) are true.[20] Here however, "$T(X_1, X_2)$" is, according to Nagel, a linguistic expression that has predicate variables in the place where the usual representation had specific theoretical terms. I have also replaced the theoretical term τ_1 in (ii) by a predicate variable. That might be somewhat unfair to what Nagel intended. Some terms in a theory may not have some experimental notion associated with them (they do not occur in the dictionary). He says "in effect those terms have the status of *variables*" (132, stress E. N.). So it looks like the occurrence of the theoretical term in (ii) should not be replaced by a predicate variable in (ii)′ since τ_1 has an experimental notion associated with it. Nevertheless, if the theoretical terms in the hypothesis part of the theory are replaced by variables, and if those same terms which may also occur in a dictionary entry are not replaced by a variable, but are kept as they are, then the deductive connection may be broken, and the explanation destroyed. Here is an example of that possibility (assuming that τ and μ are monadic theoretical terms, o and o' are experimental ones, and the universal quantifiers are suppressed):

(iv) $\tau = \mu$
(v) $o \rightarrow \tau$
(vi) $\mu \rightarrow o'$
(vii) $o \rightarrow o'$

This is a perfectly fine deduction of what might be an experimental generality (vii). However if we replace the terms in (iv) by distinct variables, but leave the occurrences of those terms untouched in the dictionary entries (v) and (vi), the deduction of (vii) is ruined.

[20] Nagel's view is subtle. He does say in effect that *if* it is required that every premise in an explanation is either true or false, then it is almost unavoidable to require that they be true (*The Structure of Science*, pp. 42–43). Factivity for him is a conditional. So, for him one could say that factivity holds (vacuously) for those premises in explanations that are, like (i)′, without any truth-value.

It is obvious that in order to retain the deduction, you either have to leave all the theoretical terms alone, or replace them all by predicate variables. Since it is a deductive explanation that is at stake, I think the thing to do is to replace all the occurrences of the theoretical terms by appropriate variables. In that case, we are left with the task of trying to explain the rules provided by (v) and (vi). The original intention of dictionary entries was to show how a theory can be related to experimental laws, but it is something of a mystery to me how that is achieved by having the dictionary give various ways in which observational terms are related to variables. Thus, if the theoretical statements have no truth-value, I think the same thing goes for the dictionary entries. And although Ernest thought of these entries as rules, I do not see how that removes the difficulty.

There is another difficulty, relatively minor, with Nagel's proposal to replace theoretical terms with variables. It treats variables in a very nonstandard way. Normally if we have a variable ranging over a certain domain, and want to consider a special case, then the usual thing is to use the name of the special element of the domain. So if one wanted to go from "*x* is a prime number other than 2, and is odd" which has no truth-value, to the special case of the number 3, one would use the name of that number to obtain "3 is a prime number and is odd." Ernest however used another convention. His example (132) is

(1) For any *x*, if *x* is an animal and *x* is *P*, then *x* is a vertebrate,

where "*P*" is a predicate variable, so that we have an expression that has no truth-value. He suggested that one way to obtain a true statement was to substitute "mammalian" for the predicate variable. The result of this "substitution" is

(2) For any *x*, if *x* is an animal and *x* is mammalian, then *x* is a vertebrate.

What Ernest says is of course true, but it is not the way substitution for variables goes. If it were a case of substitution, then what would replace the variable is the name of something in the range of that variable. In the case of predicate variables, the substitution requires the name of a predicate and not the predicate itself. Strictly speaking then, predicate variables are clearly being used in a nonstandard way. In short, Nagel's construal of theoretical terms as predicate variables is unfortunate. It threatens even his central theme that all explanations are deductive.

IX. EXPLANATION: THE SCHEMATIC ACCOUNT

What would better suit Nagel's purpose is a fresh start with the use of some device other than predicate variables to describe those premises that use theoretical terms in the explanations of laws. We suggest the systematic use of schematic letters to construct a new, but closely related account—call it N^*. It is a close cousin of Nagel's account, and would better represent his view as closely aligned with the views of writers like M. Pasch and D. Hilbert.

Schematic letters are linguistic expressions of various kinds: schematic sentential letters and schematic predicate letters, for example. Wherever they occur in expressions, they may be replaced by specific sentences and specific predicates (respectively). As we shall see, many of the difficulties connected with Nagel's replacement of theoretical terms can be resolved when they are treated schematically.

Our idea is that instead of replacing theoretical terms by predicate variables, they be replaced by schematic letters of the appropriate kind. Nothing has changed in N^*, so far as truth-values are concerned. Axioms or premises that have had their theoretical terms replaced by schematic letters will have lost whatever truth-value they might have had.

Schemas are any expressions that have schematic letters embedded in them, of whatever kind. The way to convert a schema to an expression that has a truth-value is to replace the schematic letters in them by specific nonschematic expressions of the appropriate kind, and this of course can be done in many ways. What I am suggesting then is that Nagel's view of the abstract calculi that he saw "embedded" is best represented as a schematic theory.

We have already referred to the remarks of Pasch on geometry which involved replacing terms by other terms and retaining a proof. Those remarks were sharpened by Hilbert to make the schematic features of his axiomatizations of mathematical and scientific theories more evident. It was reported by Otto Blumenthal that Hilbert, on the way back to Königsberg after hearing a lecture of geometry that was abstract (1891?), said

> One must be able to say at all times—instead of points, straight lines, and planes—tables, chairs, and beer mugs.

And in the correspondence with Frege, he said

> …you say that my concepts, e.g. 'point', 'between', are not univocally fixed.…But it is surely obvious that every theory is only a scaffolding (schema) of concepts together with their necessary connections, and that the basic elements can be thought of in any way one likes. E.g. instead of points, think of a system of love, law, chimney sweep…which satisfies all the axioms; then Pythagoras' theorem also applies to these things. Any theory can always be applied to infinitely many systems of basic elements for one only needs to apply a reversible one-one transformation and then lay it down that the axioms shall be correspondingly

> the same for the transformed things (as illustrated in the principle of duality and by my independence proofs.)...Thus the circumstance I mentioned is never a defect (but rather a tremendous advantage) of a theory.[21]

One important part of this statement is the claim that a theory is only a scaffolding (schema) of concepts. The word that was translated as "scaffolding" in the correspondence with Frege is "*Fachwerk*," and it was used very frequently by Hilbert and his students in his published articles and the now-published lectures and notes on logic and physics.

In my opinion the translation "scaffolding" is misleading. Scaffolds are constructions that are removed once a building is completed. Obviously Hilbert did not intend that the theories be cast off. "*Fachwerk*" has the standard meaning of a kind of timber-wood construction very common in Europe from the Middle Ages onward. It is a framework that is embedded in the construction, but not covered over. It was supposed to be seen from both inside and outside the building. There are in this framework various empty spaces which can be filled with stucco, brick, mortar, or adobe. It was understood that these interstices of the *Fachwerk* could be filled in many various ways, provided of course that they met the constraints imposed by the framework. The theory is supposed to remain the same, invariant, despite the various "infill" that could be used.

There are several ways to use the idioms of contemporary logic to describe what Hilbert meant by "*Fachwerk*." Other translations use "framework," and still others use model-theoretic terminology. I prefer "schema" simply because that is the word he sometimes used for "*Fachwerk*," and he does speak of "filling in the *Fachwerk*," which fits nicely with the idea of schematic letters.

Here is an example of this kind of axiomatization of Euclidean geometry provided by Hilbert and Bernays:[22]

> Axiom II.1. $(x)(y)(z)\ [Zw(x, y, z) \rightarrow Gr(x, y, z)]$, to be read as "If x lies betweeen y and z then x, y, and z lie on a straight line."

They say that the three-place predicates "*Zw*" and "*Gr*" are not the names of any specific predicates in particular. These terms are not names, but can be replaced by other three-place predicates, subject

[21] Gottlob Frege, *Philosophical and Mathematical Correspondence*, ed. Gottfried Gabriel et al. (Chicago: University Press, 1980), p. 42.

[22] David Hilbert and Paul Bernays, *Grundlagen der Mathematik*, Bd. I (Berlin, Germany: Springer, 1934), p. 6.

of course to the constraints set by the axiom. As Bernays expressed the matter,

> Thus the axiom system itself does not express something factual; rather, it presents only a possible form of a system of connections that must be investigated mathematically according to its *internal* [*innere*] properties.[23]

We can then express the schematic form of the axiom as

$$(x)(y)(z)\ [S(x, y, z) \rightarrow S^*(x, y, z)],$$

where S and S^* are predicate schematic letters. Clearly this axiom has no truth-value, and the same holds for all the axioms taken together.

Replacing some of the theoretical terms by the appropriate schematic letters avoids the problems raised by Nagel's use of predicate variables. With a shift to the schematic account N^*, several benefits are automatic. First, the sequence (iv)–(vii), with the theoretical terms replaced by schematic letters, now becomes a genuine deduction—something not true if those terms are replaced by predicate variables.[24] Second, it is impossible to make sense, say, of a dictionary entry that has the form of $o \rightarrow \tau$ (to use the made-up example of (iv)–(vii) above), if "τ" is replaced by a variable, since there is no sense provided for a conditional with a variable as consequent (or antecedent). However it is relatively easy to make sense of a conditional with a consequent (or antecedent) that is a schematic letter.

As Ernest has emphasized, the dictionary items are there to provide a deductive link to observational material. This, the shift to schematic letters achieves.

X. EXPLANATION: WHITHER CAMPBELL'S DICTIONARY?

There is however to my mind a serious oversight with Nagel's claim that if there is to be a link from a theory to observational matters then dictionary items are indispensible. It needs qualification. The claim is plausible if the link is supposed to be deductive.

However, there are two considerations under which the claim is incorrect. The first is that there is a link between some theories and experimental matters, only that link is not deductive, and not probabilistic either. There are many examples of significant theories that are regarded as having great scientific worth because of their many

[23] Bernays, "Hilbert's Significance for the Philosophy of Mathematics" (1922), trans. Paolo Mancosu, in Mancosu, *From Brouwer to Hilbert: The Debate on the Foundations of Mathematics in the 1920s* (New York: Oxford, 1998), pp. 189–97, at p. 192.

[24] According to Nagel, "If the theory is to be used as an instrument of explanation and prediction, it must somehow be linked with observable materials" (*The Structure of Science*, p. 93).

successes (the germ theory of disease, the kinetic theory of gases, and the Hamiltonian theory of least action are a few examples). These theories do not have the form of universally quantified conditionals. They have been called "theories-for,"[25] and none of the well-known experimental successes of such theories follow deductively or probabilistically from them. Moreover, no special link or coordinating definition is required to make the connection of those theories with their successes. The second is that in the present case, where the theory is understood to be schematic, there is a way in which a theory can have links to observational matters without the need to have dictionary entries. Since an axiomatized theory in general (for Hilbert) is schematic, there are many different ways in which the various schematic letters in it can be replaced by specific predicates and relations to yield statements, which unlike the schemata, have a truth-value. Let us call the result of such replacements, if true, *applications* of the theory.

Thus, to replace "point," "line," and "plane," now thought of as schematic letters, by "tables," "chairs," and "beer mugs" we do not need coordinating definitions relating these two groups of terms. It suffices to replace one set of terms for the other directly into the schematic theory, and to check whether the result is true. Hilbert himself used such replacements. He noted that by such replacements, some of the axioms of Eudoxus's theory of ratios yielded a general statement about genetics. Of course given his claim that "point" in his axiomatization of Euclidean (plane) geometry could be replaced by "ordered pair of real numbers," and "line" by "linear equation," to obtain a true application, it becomes clear why he wanted to say that ordered pairs of real numbers are also points. It is also clear that he did not mean that points could be explicitly defined as ordered pairs (or triples) of real numbers. Clearly some applications have more interest than others.[26] But none of them are deductive consequences of the schematic theory.

[25] A discussion of theories-for, their importance for understanding the nondeductive relation of them to their successes, and the kind of explanation that they provide can be found in Sidney Morgenbesser and Arnold Koslow, "Theories and Their Worth," this JOURNAL, CVII, 12 (December 2010): 616–47.

[26] In fact, Hilbert had a short proof showing that there are an infinite number of such applications, which was subsequently rediscovered by the mathematician M. H. Newman, and Donald Davidson. Hilbert wrote to Frege, "Any theory can always be applied to infinitely many systems of basic elements. For one only needs to apply a reversible one-one transformation and then lay it down that the axioms shall be correspondingly the same for the transformed things (as illustrated in the principle of duality and by my independence proofs." Reprinted in Frege, *op. cit.*, p. 42; abridged from the German edition by Brian McGuinness, ed., and Hans Kaal, trans.

I have no idea how to work out the details of Hilbert's suggestion that "points," "straight lines," and "planes" of Euclidean geometry could be replaced by "tables," "chairs," and "beer mugs" (or, as he said in his correspondence with Frege, "instead of points, think of a system of love, law, chimney sweep…which satisfies all the axioms…"). However there are other applications like the replacement of "point" by "intersection of two light rays," and "straight line" by "path of a light ray in homogeneous media," and one can see how the resulting application would be a generalization of geometrical optics, with no need for coordinating definitions.

XI. EXPLANATION, SCHEMATIC SUBSUMPTION, AND FACTIVITY

Granted that this is just a sketch, the detailed working out of this program is, we think, worth pursuing. It indicates how some laws (in the last case laws like refraction and reflection of geometrical optics) might be subsumed under a theory. We shall say that the applications are *schematically subsumed* under the theory, meaning that it is just a special case of replacement of the various schematic letters of the theory. The applications of a schematic theory are all uniformly subsumed under the same theory, and that I think counts for some type of explanation. Of course it is not the kind of subsumption endorsed by, say, Hempel when he said

> …I think that all adequate scientific models and their everyday counterparts claim or presuppose at least implicitly the deductive or inductive subsumability of whatever is to be explained under general laws or theoretical principles.[27]

Hempel was concerned with deductive and probabilistic kinds of subsumption. The applications of a schematic theory, however, are neither deductive nor inductive consequences of the theory from which they were obtained. Nevertheless, if it is possible to think of deductive (or inductive) subsumption as explanatory, then why not make the similar claim when all the applications of the theory are uniformly subsumed under it? At the very least it is also a unification.

There is still the fact that even if we go schematic with the N^* version of Nagel's original position, both accounts fail to satisfy the factivity condition for explanations of laws. Theoretical statements have no truth-value, whether, as with Nagel, it is because theoretical terms are construed as predicate variables, or whether, as in N^*, the theoretical terms are construed as schematic letters.

[27] Hempel, *Aspects of Scientific Explanation, and Other Essays in the Philosophy of Science* (New York: The Free Press, 1965), pp. 424–25.

For Ernest the distinction between experimental laws (containing no occurrences of theoretical terms) and theoretical laws (which do have one or more occurrences of theoretical terms in them) makes a difference in the issue over factivity. There are a few possibilities worth mentioning: (1) An experimental law is explained by other experimental laws. In this case factivity holds. (2) In the case of an explanation of an experimental law by theoretical ones, matters are different. Some schematic premises deductively yield and perhaps even explain a nonschematic generalization. One example of this possibility might be the argument (iv) to (vii) above. Here the experimental law (vii) is true (say) but what explains it is schematic, and so without truth-value. In such a case we might say that the explanation is *semi-factive*, a condition which requires only that what is explained is true (rather than what explains must be true). In this case the situation is comparable to that of the factive "knows that…". (3) Some theoretical laws explain other theoretical laws. In this case there is no factivity at all, not even semi-factivity, since the explainers and the explained do not have truth-values. This may raise problems for Ernest's theory of reduction since he seems to regard theory reduction as a case of explanation of one theory by another.[28] Maybe not. Recall that for Ernest, factivity is conditional: if the items in an explanation have any truth-value at all, then they are true. So in this case, factivity would be satisfied vacuously!

There is yet another way in which factivity is involved even with theories that are schematic. Such theories we have said have their applications (obtained by replacement of their schematic letters with appropriate predicates and relations). Those applications of a theory will in general not be schematic, and will in turn figure in the explanation of other nonschematic statements. For those explanations that are based on applications that are subsumed under the theory, nonvacuous factivity still prevails.

XII. EXPLANATION AND GENERALITY

We turn finally to a problem which both Braithwaite and Nagel recognized as an important, maybe even a crucial element of the explanation of laws. It seems that there is no explanation of a law which does not involve at least one premise that is more general than it. As far as I can tell, only Nagel attempted a formal account of the concept

[28] "Reduction…is the explanation of a theory or a set of experimental laws established in one area of inquiry, by a theory usually but not invariably formulated for some other domain." Nagel, *The Structure of Science*, p. 338.

"is more general than." Unfortunately, I do not see at present how his account fits into our present (schematic) account of Nagel's view of the explanation of laws. There are several drawbacks to his account of generality.

First there is the restriction of his analysis to only those laws that are best expressed as universally quantified conditionals. There are, as we have noted, some laws and theories of high scientific merit which do not have that elementary form. Second, it appears, on his account, that whether one law is more general than another depends critically on syntactic form, so that the order of generality between two laws can change if one uses different, but logically equivalent forms of those laws. And finally, it seems that the account is incompatible with simple examples of explanations offered by Nagel.

Ernest has two accounts of generality: one which addresses the problem of when one group of laws (say, physics) is more general than another (say, a group of biological laws), and a simpler version for comparing one law with another. We concentrate on the simpler version. It goes like this: Let *L* be the law that All *A*s are *B*, and *L** be the law that All *C*s are *D*. Then,

> *L* is more general than *L** if and only if it is logically true that All *C*s are *A*, and it is not logically true that All *A*s are *C*. (38)

I do not find it so damaging that the laws for which this concept is defined are limited to generalized conditionals. It would still be an important achievement even so. The second feature of this definition is that two logically equivalent expressions of a law can differ in generality. That seems to me not only surprising, but unwelcome. Ernest was well aware of this feature. His example is that

> (1) "All living organisms are mortal" is more general than
> (2) "All human beings are mortal" on his account, but
> (3) "All nonmortals are nonliving organisms"

is not more general than (2) "All human beings are mortal," even though (1) and (3) are logically equivalent.

Ernest has an interesting defense to mitigate this result. It depends upon his deep commitment to a pragmatic view of the matter. He thinks that there is "a tacit reference to contexts of use in the formulation of laws" in which the range of application of the law (indicated by the antecedent of the conditional) can shift. His example is that the common use of "Ice floats in water" has as its range of application cases of ice that are, were, or will be immersed in water, and he says that rarely (if ever) will the range of application be those things that never float in water. He may be right in this observation, but I do not see what the target of investigation (what happens to ice immersed in water) has to do with the generality of a statement. That matter may be moot, but there is something more serious at stake. This result leaves open the possibility that you may lose an explanation of a law by using one formulation of a law rather than its logical equivalent. Here is the possibility that I have in mind. Suppose that there is an explanation of the form (i):

(i) $A \rightarrow B$, $[\neg C \rightarrow (B \rightarrow \neg A)]$ implies $(A \rightarrow C)$. Now,
(ii) $[\neg C \rightarrow (B \rightarrow \neg A)]$ is logically equivalent to $[(A \wedge B) \rightarrow C]$, therefore
(iii) $A \rightarrow B$, $[(A \wedge B) \rightarrow C]$ implies $(A \rightarrow C)$,

but (iii) is not an explanation. The reason is that neither of the premises has greater generality than the conclusion. If $A \rightarrow B$ has greater generality than $(A \rightarrow C)$, then $(A \rightarrow A)$ is logically necessary, and it is also not logically necessary. Clearly impossible. And if $[(A \wedge B) \rightarrow C]$ has greater generality than $(A \rightarrow C)$, then $[A \rightarrow (A \wedge B)]$ is logically necessary, and $[(A \wedge B) \rightarrow A]$ is not logically necessary. Also clearly impossible. Consequently, there can be a loss in explanations with the substitution of logical equivalents.

The most serious drawback to this account of "is more general than" is that clear examples of explanations of laws fail to meet the condition that at least one of the premises has greater generality than the law to be explained. The example is Nagel's. He offers a specific example of an explanation of the law, "When gases containing water vapor are sufficiently expanded without changing their heat content, the vapor condenses." It has the form

(T) All *A*s are *C*s and All *C*s are *B*s, therefore All *A*s are *B*s.

Now "All *A*s are *C*s" cannot be more general than "All *A*s are *B*s," for then we would have that "All *A*s are *A*s" is a logical truth, and also that it is not. Clearly impossible. Moreover, if "All *C*s are *B*s" is more general than "All *A*s are *B*s," then "All *A*s are *C*s" is logically necessary. But that is one of the two premises of the explanation, and Nagel's view is that scientific laws are not logically necessary (52–56). The conclusion is therefore that there are no explanations of laws that have the form of (T)—not even the nice one that he provided.

Despite the fact that this specific proposal for explicating the variation in generality among laws does not work, it should be possible with some further effort to fit this notion into the schematic framework we have attributed to Nagel.[29]

Norman Campbell provided an initial source for the work of Nagel and Braithwaite, with his canonical representation of theories. They incorporated that structure into their accounts of laws and explanations in science. Campbell's assumption that all scientific explanations involved deductions was retained by both Nagel and Braithwaite. From that point on however, the differences widened under the influence of two different traditions, to yield two remarkable original, viable accounts. Campbell regarded the hypothesis and dictionary entries as statements with truth-values. That assumption was retained by Braithwaite, but was rejected by Nagel. Nagel, in contrast, followed a tradition defended by the mathematician D. Hilbert, who developed the idea that theories were neither true nor false. Ernest thought theories were statement-forms without truth-value. Under the influence of J. S. Mill and Ramsey, Braithwaite developed a notion of law indexed to special logical structures which he called deductive systems, but the notion of a law was not tied by Nagel to membership in any specially dedicated deductive structure. Lastly it is to Ernest that we owe an account of "is more general than" which plays a prominent role in explanations of laws: some of the laws that explain others must be more general than them. Braithwaite seems to have thought that the laws of his established deductive systems were ranked by generality, but he offered no account of that central idea.

ARNOLD KOSLOW

City University of New York Graduate Center

[29] I am not too optimistic about the prospects of an account that depends so crucially on the scope of attribution. That concept makes sense only when the laws are universally quantified conditionals. Most of the Hilbert and Bernays axioms for Euclidean geometry (as cited above) are not conditional in form. But considerations of schematic subsumption would not cover the convincing examples of nonschematic examples of comparative generality that Nagel has provided. It is still an open problem as far as I can see.

REFLECTIONS ON ERNEST NAGEL'S 1977 DEWEY LECTURES *TELEOLOGY REVISITED*

It is now over 60 years since I took my first graduate course in philosophy at Columbia University in January of 1947. The course was a seminar by Ernest Nagel on the philosophy of logic of Bradley and Dewey, and by the end of the second meeting I was enthralled by Ernest's patient style and careful criticisms of their logical views. It is still easy for me to visualize Ernest as he walked back and forth in the seminar room complimenting, on the one hand, the philosopher being studied, but always ending, on the other hand, on a clear statement of something that was wrong in the text being examined. As I have been reading, in preparation for writing this article, what he wrote about biology, mostly after I left Columbia in 1950 to come to Stanford, the cadence of the written words and the style of argument remind me very much of those first days of listening to him in 1947. Yet in the years I was at Columbia, Ernest did not give a single seminar or course on the philosophy of biology, nor did anyone else at Columbia that I can remember.

It is certainly different now. There are more biologists at Stanford than any other kind of scientist, by a wide margin. Courses in biology are now given everywhere and for every kind of purpose. Stanford, for example, now has a separate major in human biology. And most philosophy departments now feel obligated to give at least one course in the philosophy of biology. Not surprisingly, Ernest, with his wide scientific interests, began earlier than most philosophers to think about biology.

Nagel's first article I know about that has an interesting bearing on biology is a very useful one on naturalism and materialism.[1] From the style of the arguments, I would judge that the article was mainly written by Nagel, the youngest of the three authors. After this there are two important articles just by Nagel himself.[2] Then there is a break until 1961 with the publication of his magisterial book.[3] Several

[1] John Dewey, Sidney Hook, and Ernest Nagel, "Are Naturalists Materialists?" this JOURNAL, XLII, 19 (Sep. 13, 1945): 515–30.

[2] Nagel,"Mechanistic Explanation and Organismic Biology," *Philosophy and Phenomenological Research*, XI, 3 (March 1951): 327–38; Nagel, "Wholes, Sums, and Organic Unities," *Philosophical Studies*, III, 2 (February 1952): 17–32.

[3] Nagel, *The Structure of Science: Problems in the Logic of Scientific Explanation* (New York: Harcourt, Brace and World, 1961).

0022-362X/12/0908-09/503–15

parts of it deal with biology, and I shall return to discuss them later. After another break, he published two long articles containing his John Dewey Lectures at Columbia in March of 1977, together titled *Teleology Revisited*, which appeared almost immediately in this JOURNAL,[4] and were later reprinted in a volume of essays.[5]

In the rest of this article, I will devote most of the time to analyzing the parts of the 1961 book *The Structure of Science* devoted to biology, and the 1977 John Dewey Lectures on teleology. *The Structure of Science* is densely argued from chapter to chapter as it surveys the natural and social sciences. Chapter 12 is entirely devoted to biology. Its main focus concerns the long controversy about teleological explanations. It was my original intent to survey rather carefully everything Nagel has written on the philosophy of biology, but the more I re-read his published work of half a century ago, I came to see that there were too many different topics to discuss them all with any thoroughness. Moreover, it seems to me that his analysis in the book of teleological explanations, and his return to this topic in his 1977 Dewey Lectures, represented his most serious and engaged work in the philosophy of biology. At the same time, the tangled issues he examined with care remain important.

So, I have organized this article into five sections. The first two deal with teleological explanations, first, in the 1961 book, and second, in the first Dewey Lecture. The third section, representing the second half of chapter 12, focuses on the stance of organismic biologists toward mechanistic explanations. The fourth section examines Nagel's extended analysis in the second Dewey Lecture on functional explanations in biology. Finally, the fifth section consists of my own comments on some of the contemporary issues about the nature of biology as a science, related to Nagel's critical analysis.

I

It is not possible to give a reasonable paraphrase of the many balanced arguments Nagel gives in chapter 12 of *The Structure of Science* on the status of teleological explanations. I mention one point that he covers at the beginning in rather short order, but that is a topic of much greater concern currently. This is the status of *conscious* purposes. I am sure this is because the status of consciousness as such did not occupy anything like its current prominence when he

[4] Nagel, "Goal-Directed Processes in Biology" and "Functional Explanations in Biology," this JOURNAL, LXXIV, 5 (May 1977): 261–79 and 280–301.

[5] Nagel, *Teleology Revisited and Other Essays in the Philosophy and History of Science* (New York: Columbia, 1979).

was writing. He does emphasize that the conscious aspect of purpose, so to speak, is not relevant to his focus on teleological explanations.

He does give much more attention to the claim often made in the past, and still by some, that teleological explanations are an essential part of biology, as a special subject matter for them. He has a number of arguments against this position, among them that we can, correctly, use extremal principles to formulate classical mechanics in teleological language. Another closely related one is the search for stability results in mechanics, which resembles arguments concerning the important goal of persistence in biology.

A related point, but from a different angle, is the claim that the special character of biology is that it studies organic systems, which are fundamentally goal-directed. Nagel responded to this claim by asking, first, can we, in fact, identify the particular structures of goal-directed systems in a neutral way? And, second, does the current focus of biology constitute in itself strong evidence for this claim? I cannot resist quoting the last part of his response to these questions.

> Moreover, some systems have been classified as "teleological" at one time and in relation to one body of knowledge, only to be re-classified as "nonteleological" at a later time, as knowledge concerning the physics of mechanisms improved. "Nature does nothing in vain" was a maxim commonly accepted in pre-Newtonian physics, and on the basis of the doctrine of "natural places" even the descent of bodies and the ascent of smoke were regarded as goal-directed. Accordingly, it is at least an open question whether the current distinction between systems that are goal-directed and those that are not invariably has an identifiable objective basis (i.e., in terms of differences between the actual organizations of such systems), and whether the *same* system may not often be classified in alternative ways depending on the perspective from which it is viewed and on the antecedent assumptions adopted for analyzing its structure.[6]

Later, I add to this passage some engineering examples of my own. Nagel then goes on to endorse as the most reasonable approach the system approach to biology, exemplified, at the time he was writing, by G. Sommerhoff,[7] and the work of Norbert Wiener and colleagues.[8]

To the exposition and defense of this system view, Nagel gives an especially careful presentation, because he endorses it, more than

[6] Nagel, *The Structure of Science*, p. 419.

[7] G. Sommerhoff, *Analytical Biology* (London, UK: Oxford, 1950).

[8] Arturo Rosenblueth, Norbert Wiener, and Julian Bigelow, "Behavior, Purpose and Teleology," *Philosophy of Science*, x, 1 (January 1943): 18–24; Wiener, *Cybernetics: Control and Communication in the Animal and the Machine* (New York: Wiley, 1948).

any other, in its effective way of generalizing the argument away from purely biological subject matter to other scientific structures, physical as well as biological, where, here, "physical" means in the sense of physics. His defense was strengthened by the simultaneous appearance of a number of important works on this topic. The significant detailed points that he makes are too numerous to enumerate here, so I end by quoting his summarizing final paragraph.

> It follows from these various considerations that the distinction between structure and function covers nothing that distinguishes biology from the physical sciences, or that necessitates the use in biology of a distinctive logic of explanation. It has not been the aim of the present discussion to deny the patent differences between biology and other natural sciences with respect to the role played by functional analyses. Nor has it been its aim to cast doubt on the legitimacy of such explanations in any domain in which they are appropriate because of the special character of the systems investigated. The objective of the discussion has been to show only that the prevalence of teleological explanations in biology does not constitute a pattern of explanation incomparably different from those current in the physical sciences, and that the use of such explanations in biology is not a sufficient reason for maintaining that this discipline requires a radically distinctive logic of inquiry.[9]

II

I now pick up his analysis of teleological explanations sixteen years later. In direct reference to earlier work, Nagel says at the beginning of the first Dewey Lecture the following: "The recent literature sometimes breaks fresh ground; but much of it consists of modified (and sometimes much improved) versions of well-known views, presentations of difficulties in older analyses, or challenges to basic assumptions underlying customary approaches. My discussion will, in consequence, inevitably revisit much familiar territory, though with some different objectives than on previous journeys."[10] He then proceeds to introduce the analysis of several different views of teleology he did not cover in the book, most of these being generated by publications that appeared later.

The first new view he considers is the intentional one, which emphasizes human purposes and goals as primary. Following this idea, Nagel refers in schematic form to the belief-desire model, according to which goals can be achieved by alternative methods internally evaluated. In my view, to generalize this intentional human

[9] Nagel, *The Structure of Science*, p. 428.
[10] Nagel, "Goal-Directed Processes in Biology," p. 262.

model, formal similarity or isomorphism to human behavior should be found. Such investigations are familiar in other parts of biology, as can be seen in the detailed anatomical studies of the structure of eyes in various species. On the other hand, I am skeptical of the belief-desire model as being appropriate. It is too far from a psychologically plausible model, as I have argued elsewhere.[11]

The second view considered is the one that Nagel says is derived from information theory, and he labels it the *program* view. The discussion expands upon the earlier one about programming described in the 1961 chapter, but is now revised as a reflection, I am sure, of the rapid development of computer science and the widespread use of computers between 1961 and 1977. Perhaps his most significant example is the one made prominent by the DNA discoveries about the programming of the human genome. For more detailed considerations, he examines the approach of the prominent biologist Ernst Mayr.[12] Within teleology, Mayr introduces the distinction between processes that are teleomatic, or in other words, automatic, as is the case for many human habits, and, on the other hand, "teleonomic," which is close to many standard uses of the word "teleological."

Nagel has three principal objections to Mayr's view. The first is that this concept is not normally used in the definition of goal-directed processes, and does not seem essential. His second objection is that control of a process by a program is not sufficient for the process to be judged goal-directed. His counterexample to show this is the knee-jerk reflex, which does not have the property of persistence required of a goal-directed process. His third objection is to Mayr's distinction between open and closed programs. The intuitive idea, Nagel asserts, is clear, at least in some general sort of way, but it is hard, if not impossible, to apply the distinction in practice. (I omit the details of this argument.) Nagel's closing point is that in summary, it is also hard in general to distinguish between teleomatic and teleonomic processes.

Nagel then returns to the system view much discussed in the 1961 analysis, but now called the "system-property view." He focuses on a serious problem that was not discussed in 1961. This is the problem

[11] Patrick Suppes, "Rationality, habits and freedom," in Nicola Dimitri, Marcello Basili, and Itzhak Gilboa, eds., *Cognitive Processes and Economic Behaviour* (New York: Routledge, 2003), pp. 137–67; Aimee Drolet and Suppes, "The good and the bad, the true and the false," in Maria Carla Galavotti, Roberto Scazzieri, and Suppes, eds., *Reasoning, Rationality, and Probability* (Stanford: CSLI, 2008), pp. 13–35.

[12] Ernst Mayr, "Teleological and Teleonomic: A New Analysis," in R. S. Cohen and Marx Wartofsky, eds., *Methodological and Historical Essays in the Natural and Social Sciences* (Boston: Reidel, 1974), p. 98.

of having variables that are not determinantly connected by known laws of nature. Such independence of variables is required in the system-property view. Nagel supports it and does not find the relativistic formulation unacceptable, but an honest admission of a common state of science. Given this hypothesis of at least two variables that are appropriately independent, Nagel judges that such explanations of goal-directed processes in biology can have a structure which is like that of nonteleological explanations in the physical sciences.

Nagel concludes that the sort of explanations in terms of purposive behavior is characteristic of biology and can be recognized as being scientifically appropriate when formulated in a way that satisfies the various constraints discussed. I note that I have omitted necessarily the essential content of some of his elaborate arguments, but I believe my summary is faithful to his general position.

III

In the second half of chapter 12 of *The Structure of Science,* Nagel did not write something directly on functional explanations in biology. What he did was to write a prelude on the standpoint of organismic biologists that he defends as being the proper scientific attitude toward biology. This is the attitude that biology neither has vitalistic forces, in the sense of such nineteenth-century scientists as Driesch, nor is reducible in the other direction to physics. The second half of chapter 12 explains the positive features of this viewpoint. Now that vitalism is dead, Nagel emphasizes, at the beginning, the unifying theme among organismic biologists is this rejection of "mechanistic" explanations in the spirit of physics. And, as he also emphasizes, such biologists are often not clear about what they are objecting to. It is to these arguments that I now turn.

First, it is evident that there are areas of biology that do not need or seem to require mechanisms in their formulations. Two examples that are given are the theory of evolution and the gene theory of heredity. Second, biological processes are not reducible to the concepts and laws of physics and chemistry. Claims about holism and the strong dependence of the parts of biological systems are used to support this argument. Third, biological systems are hierarchically organized. There seems to be no comparison made here with the hierarchical physical organization of electrons, atoms, and molecules.

In reply to these arguments, Nagel points out that it is widely recognized that there is no real possibility of the full reduction of biology to physics and chemistry. Since such arguments against this reduction are well known, many of them cited by Nagel in footnotes, I will not elaborate on a point that is nonetheless important. Nagel stresses the openness of the extent to which physical and chemical laws can be used to explain biological phenomena. It is fair to say that writing half a century after the appearance of *The Structure of Science*, I find the use of physical concepts and arguments in biology is much greater, but any complete reduction of biology to purely physical ideas remains out of sight. Indeed, the famous remark of Dirac that, with the discovery of quantum mechanics, chemistry has become an applied branch of physics, represents an equally unattainable ideal in the framework of modern chemistry, which is flourishing as much as ever as a discipline separate from physics.

Nagel ends the chapter by saying that the main conclusion is that organismic biologists have not established the absolute autonomy of biology as a scientific discipline. The ironic comment to be made these many years later is that something else that would have seemed very surprising in 1961 is that biology is in some sense absorbing the other sciences. For example, many of the doctoral dissertations at Stanford in applied physics, electrical engineering, and computer science are now about some biological topic. Moreover, the School of Engineering at Stanford has many undergraduates, but now a large number are in the new Department of Bioengineering. From a research standpoint, literally thousands of papers are being produced each month worldwide on biological experiments that use increasingly sophisticated physical apparatus. What is happening is the intimate and far-reaching commingling of the physical and biological sciences, which suggests that in the future the relevant question will not be that of the reduction of biology to physics and chemistry, but rather, can we actually distinguish and separate these disciplines in a meaningful way.

IV

In several ways the second John Dewey Lecture is of greater general importance to Nagel's philosophy of biology than the first. In discussing this topic, I follow the pagination of the version published in this JOURNAL, which is essentially the same as what was published two years later in the volume already referred to. The initial and the longest part of the article concerns Nagel's careful and extended survey of the variety of views on functional explanation that may be found in the literature circa the 1970s or somewhat earlier. To structure my exposition explicitly, I number each of these views in the order they are set forth in the article. To begin with, here are two of Nagel's simple examples of functional explanations: "fish have gills in order to obtain oxygen" and "human blood contains leucocytes for the sake of defending the body against invading bacteria."[13]

(i) The first type of functional explanation is the neutral view, as Nagel puts it, "teleologically neutral." An example to illustrate this idea is given from physics—the function of mass in the motion of a physical object. Nagel notes that in physics such functions have the same role as explanations, which in physics have no teleological connotations. He cites as an example of such a neutral view in biology the functional explanations given in an article by Walter Bock and Gerd von Wahlert, who essentially equate biological functions with biological roles.[14] It is this treatment of roles as functions and nothing more that makes their view a neutral one. They define the biological role of a faculty in an organism as "the action or use of the faculty by the organism in the course of its life history," and I quote Nagel, who partially quotes them: "…the biological role of the large mucus glands of gray jays is said to be the use of those glands 'as a glue to cement food particles together into a food bolus which is then stuck to the branches of trees'."[15] Nagel is more or less accepting of this kind of functional explanation, as one that works similarly in physics, where Nature's plan or the role of complex numbers in quantum states are often referred to, but in an innocent metaphorical way. I take it the concept of role in this neutral view is so being interpreted by Nagel.

(ii) His main criticisms are devoted to the widespread second view of functional explanation in biology. This is the one that mainly assumes that biological functional explanations are teleological, in the sense that they are about actions directed by some agent who has a goal in mind. Nagel focuses on the discussion of this view by Larry Wright, who strengthens the account of selective agency to require an analogy to some "conscious" function in human behavior.[16] Put another way, Wright holds that functional explanations in biology should have the same pattern as do explanations of functions of which humans are conscious. Nagel rightly offers strong criticism of this view, which as far as I know does not have much current attention. Among Nagel's criticisms I mention his argument that natural selection in evolution is selection by chance, as opposed to selection

[13] Nagel, "Functional Explanations in Biology," p. 280.

[14] Walter J. Bock and Gerd von Wahlert, "Adaptation and the Form-Function Complex," *Evolution*, XIX, 3 (September 1965): 269–99, at p. 274.

[15] Nagel, "Functional Explanations in Biology," p. 281.

[16] Larry Wright, *Teleological Explanations: An Etiological Analysis of Goals and Functions* (Berkeley: California UP, 1976).

by a conscious agent. Nagel points out that the attempt to make Darwinian selection in any sense agency-oriented is a mistake, even though many biologists and more theologians have long yearned to show that such an agency view is plausible.

(iii) This view gives functions a purely methodological or heuristic role. As Nagel remarks, it is close to Kant's complicated formulation of teleological principles as regulative, rather than constitutive, for guiding biological and human inquiry. Kant, of course, had to struggle to make these ideas consistent within the framework of universal mechanical laws. This is why the teleological principles are not constitutive of nature, but maxims. I note that a surprisingly similar view is adopted widely in the use of statistics, and probability theory more generally, in the classical physics of the nineteenth and twentieth century. The conceptual conflict that seems apparent is dealt with in a way that surely would have satisfied Kant. For more discussion of this matter, see my article.[17]

(iv) The fourth view, which Nagel regards as more plausible than the three discussed so far, is that such functions contribute to the "welfare" of either individual organisms or populations of them. He first examines the critique of Carl Hempel, a well-known philosopher who was also a good friend of his, so his treatment of this view is sympathetic, even though critical. The heart of Hempel's critique is that a functional explanation must provide a specific explanation to be satisfactory, but if the occurrence of any one of a number of alternatives would work as well, then no functional explanation of the presence of some specific cause has been given. The issue Hempel's analysis generates is associated with the doctrine of the plurality of causes, as Nagel points out. Yet a possible plurality of causes, or in Hempel's language, possible alternatives, is not sufficient to invalidate a detailed argument. Nagel uses the familiar example of determining the cause of death, which is in fact often straightforward, even though there are cases in which a plurality of causes is a real problem. In my version of Nagel's critique of Hempel's analysis I have much simplified the argument to what I judge is the main point.

Nagel moves on in his exposition to the version of the welfare view proposed by Michael Ruse.[18] The special interest here is that Ruse gives his analysis as part of a critique of some of Nagel's own views.

[17] Suppes, "Indeterminism or Instability, Does it Matter?" in Gordon G. Brittan, Jr., ed., *Causality, Method, and Modality: Essays in Honor of Jules Vuillemin* (Boston: Kluwer Academic, 1991), pp. 5–22.

[18] Michael Ruse, *The Philosophy of Biology* (London, UK: Hutchinson, 1973), chapter 9.

In responding to Ruse's criticisms, Nagel also makes his own criticisms of Ruse. The discussion here is extended and rather complicated. I shall try to give a sense of it, at least in a brief way, by citing one of the amusing examples. Ruse claims that the goal-seeking system to which Nagel's own formulation applies can be "altogether inappropriate." Ruse's example is this. Suppose it were true that long hair on dogs harbors fleas, and also that dogs are goal-directed toward survival. Ruse then thinks that Nagel is committed to saying that a biological function of long hair on dogs is to harbor fleas.

The essence of Nagel's response is that functional explanations presuppose not only that a given system is goal-directed, but also that the biological functions attributed to it actually contribute to the maintenance of the system. So Nagel would be committed to Ruse's claim about fleas only if he knew that in fact the presence of fleas provided some kind of immunity from disease or something similar.

Nagel does admit that there seems to be a problem if every effect of a feature will have to count as one of an organism's functions, if such effect contributes to the maintenance of some goal or other. It seems to me that Nagel is too pessimistic on this point. As scientific disciplines develop, the range of effects considered is highly restricted, as part of the ongoing scientific work. For example, there is good reason in the continued development of classical mechanics to not investigate the complicated history of most physical objects or processes being studied, except for a highly restricted list of properties.

Moreover, this restricted list can be sharply constrained either by direct theoretical argument for such an established theory, or by examining current experimental work, as is done for many subtle problems of continuum mechanics, such as the pattern of airflow around airplane wings.

Nagel ends up adopting the goal-supporting view of functional explanations in biology, yet, in his characteristic fashion, with some substantive reservations. His broad conclusions at the end of the second Dewey Lecture are these. First, when the system-property account of goal-directed processes is sound, functional explanations can be shown to have the same general structure as explanations in the physical sciences. Second, such accounts of goal ascriptions are also causal in nature. Third, functional ascriptions have a different character. They are not necessarily causal, but can also be explanations of the effects of various features and processes. Both kinds of inquiries, those about causes and those about effects, have their proper place. Finally, none of the analysis and conclusions given show that the laws and theories of biology are reducible to those of the physical sciences, but do have a proper scientific status of their own.

V

Here are some final comments of my own. First, there is one surprising and important omission in the discussion of biological explanations. It is the extensive use in biology of probabilistic and statistical concepts that leads directly to the claim that a wide variety of biological phenomena have, in the best current science, probabilistic explanations based on the probabilistic nature of many biological causes. This is not a recent development, but something that rapidly developed, at the end of the nineteenth and the beginning of the twentieth century, in the work of Galton, Pearson, and other distinguished scientists. Within this framework, many functional explanations in biology are formally stated statistically. It was a great triumph of the twentieth century to develop formal statistical tests of the correctness of such hypotheses, or even more, tests for when to reject them. Moreover, the qualitative consideration of probabilistic causes goes back to antiquity, and already in the eighteenth century, Laplace's memoir of 1774[19] developed systematic methods for inferring probabilistic causes from effects. In many ways, for much of the twentieth century, the most formal part of biology was its statistical side, and many statistical studies were normative in Nagel's sense of goal-directed. This is not the place to elaborate on these ideas, but I believe they would strengthen Nagel's claims about the extent to which biology is not reducible in any coherent way to current physics and chemistry, notwithstanding the great importance of physical and chemical mechanisms in modern molecular and cell biology.

Second, there are two striking cases of the rich use of teleological concepts in the context of the physical sciences and their engineering applications. The first is engineering design, especially in the modern subject of operational research and control theory, since the middle of the twentieth century. The extensive use of physical theory would be expected, but the surprise is the richness of the mathematical methods that have been developed in this context to serve teleological goals. These methods, such as linear programming, and later, nonlinear dynamic programming, have made possible quantitative solutions of normative problems of optimization of engineering, economic, and social importance.

Third, the idea of robots goes back to antiquity, but it is really in the twenty-first century that we will find them everywhere. Their

[19] Pierre-Simon Laplace, "Mémoire sur la probabilité des causes par les évènemens," *Mémoires de mathématique et de physique, presentés à l'Académie Royale des Sciences, par divers Sçavans* (often cited as *Savants étrangers*), VI (1774): 621–56.

materials of construction will mostly be far from biological in nature, but their principles of operation will be purposeful and clearly teleological in nature. Indeed, certain robots will be built to model human performance as closely as possible. Perhaps the most sophisticated examples of this kind will be robots that seem to talk and think in a serious but natural conversational style, including the subtle expression of emotion in speech, a subject already being studied in computer science. The modern irony of the "machine" will be that the closest friends of humans will in many cases be their robot companions, who can communicate so much better than their animal friends of long tradition. In many ways, biologically congenial teleological concepts will invade the physical and engineering sciences just as thoroughly as physics has invaded biology.

Fourth, the current obsession with gaining a scientific understanding of consciousness will surely be resolved in the present century. This inevitable achievement of neuroscience will then be used to make our robots of the future ever more human. And the scientific literature will see a welter of articles on the physics of consciousness.

Fifth, it is natural to ask how physics will have anything to say about consciousness. The answer will come from solving another problem of even greater general importance. This is the problem of understanding how the brain computes in any of its main systems tasks such as speaking or listening, coordinating accurately perception and motor control of action, or recognizing the expression of subtle varieties of emotion, present in human interactions of all sort.

There are many reasons to think that in the next few decades we will reach a much deeper understanding of how the brain computes, even if we may still be baffled by how it learns to do so. The beginning is already several decades old, starting with the still-older recognition that much biological behavior is oscillatory in nature, from the chirping of crickets and the flashing of fire-flies, to the rhythms of the heart and of musical instruments. The obvious physical devices for generating oscillations are oscillators, which are used to model an extraordinary variety of phenomena in physics from clock pendulums to quantum fields. A widely accepted hypothesis in current neuroscience is that collections of synchronized neurons in the brain form electromagnetic oscillators. Roughly speaking, there are two ways such oscillators can interact. One is by resonance, transferring energy at a common frequency from one oscillator to another, as in the famous Tacoma, Washington, bridge collapse on November 7, 1940. Enormous oscillatory energy was transferred to the bridge from the strong winds at the fundamental frequency of the bridge's main span. Such resonance is, among other things, much too slow for brain computing. So the plausible hypothesis is that the much faster phase-locking of oscillators is the main engine of system computation in the brain. Little energy is required for phase-locking, and after oscillators are synchronized in phase, no external energy is required. Weak coupling between the oscillators will suffice for phase-locking. I and my coworkers believe that an important example of such phase-locking is the mechanism of fast phase-locking for retrieving words from verbal memory in speaking or listening, as well as in writing or reading.[20] A deeper intertwining of the mental and the physical is hard to imagine. By the time these problems are completely solved, biology and physics may well be indistinguishable. I believe this is an outcome that fits in well with Nagel's elaborate analysis of the structure of science, properly revised from time to time to accommodate the science of today and tomorrow.

PATRICK SUPPES

Stanford University

[20] Suppes, "A Revised Agenda for Philosophy of Mind (and Brain)," in Michael Frauchiger and Wilhelm K. Essler, eds., *Representation, Evidence, and Justification: Themes from Suppes* (Frankfurt, Germany: Ontos Verlag, 2008), pp. 19–50; Suppes, Marcos Perreau-Guimaraes, and Dik Kin Wong, "Partial Orders of Similarity Differences Invariant between EEG-Recorded Brain and Perceptual Representations of Language," *Neural Computation*, XXI, 11 (November 2009): 3228–69; Suppes and J. Acacio de Barros, "Quantum Mechanics and the Brain," *Quantum Interaction: Papers from the 2007 AAAI Spring Symposium, Technical Report SS-07-08* (Menlo Park, CA: AAAI, 2007), pp. 75–82; de Barros and Suppes, "Quantum mechanics, interference, and the brain," *Journal of Mathematical Psychology*, LIII, 5 (October 2009): 306–13; E. Vassilieva, G. Pinto, de Barros, and Suppes, "Learning Pattern Recognition Through Quasi-Synchronization of Phase Oscillators," *IEEE Transactions on Neural Networks*, XXII, 1 (2011): 84–95; Suppes, de Barros, and G. Oas, "Phase-oscillator computations as neural models of stimulus-response conditioning and response selection," *Journal of Mathematical Psychology*, LVI, 2 (April 2012): 95–117.

WHAT KIND OF UNCERTAINTY IS THAT? USING PERSONAL PROBABILITY FOR EXPRESSING ONE'S THINKING ABOUT LOGICAL AND MATHEMATICAL PROPOSITIONS*

> What is essential for the future development of probability considerations, as for the development of science in general, is that trained minds play upon its problems freely and that those engaged in discussing them illustrate in their own procedure the characteristic temper of scientific inquiry—to claim no infallibility and to exempt no proposed solution of a problem from intense criticism. Such a policy has borne precious fruit in the past, and it is reasonable to expect that it will continue to do so.
> —Ernest Nagel, *Principles of the Theory of Probability*, Concluding Remarks[1]

Try to use probability to formalize your uncertainty about logical or mathematical assertions. What is the challenge?

Concerning the normative theory of personal probability, in a frank presentation titled *Difficulties in the theory of personal probability*,[2] L. J. Savage writes,

> The analysis should be careful not to prove too much; for some departures from theory are inevitable, and some even laudable. For example, a person required to risk money on a remote digit of π would, in order to comply fully with the theory, have to compute that digit, though this would really be wasteful if the cost of computation were more than the prize involved. For the postulates of the theory imply that you should behave in accordance with the logical implication of all that you know. Is it possible to improve the theory in this respect, making allowance within it for the cost of thinking, or would that

* We thank Jessi Cisewski and Rafael Stern for their helpful comments with prior drafts of this paper.

[1] Ernest Nagel, *Principles of the Theory of Probability* (Chicago: University Press, 1939), pp. 76–77.

[2] This text is taken from a draft of Leonard J. Savage's manuscript, *Difficulties in the theory of personal probability*, dated April 1, 1967. Savage gave one of us (JBK) this draft while both were members of the Statistics faculty at Yale University. This text agrees with the quotation on p. 311 of Ian Hacking, "Slightly More Realistic Personal Probability," *Philosophy of Science*, xxxiv, 4 (December 1967): 311–25. In the published version, Savage, "Difficulties in the Theory of Personal Probability," *Philosophy of Science*, xxxiv, 4 (December 1967): 305–10, this text appears (p. 308) with printing errors, which are duplicated also in the version appearing in *The Writings of Leonard Jimmie Savage: A Memorial Selection* (Washington, DC: American Statistical Association and the Institute of Mathematical Statistics, 1981), p. 511.

0022-362X/12/0908-09/516–33

> entail paradox, as I am inclined to believe but unable to demonstrate? If the remedy is not in changing the theory but rather in the way in which we are to attempt to use it, clarification is still to be desired.

But why does Savage assert that "a person required to risk money on a remote digit of π would, in order to comply fully with the theory, have to compute that digit"? His short answer is that the postulates of the theory of personal probability "imply that you should behave in accordance with the logical implication of all that you know."

In this essay we discuss three strategies for addressing Savage's challenge:

(1) Adapt I. J. Good's[3] idea to use a *Statistican's Stooge* in order to change the object of uncertainty for the agent. The *Stooge* replaces the problematic constant π by a nonproblematic random variable θ that the *Stooge* knows is co-extensive with π. From the perspective of the *Statistician*, the theory of personal probability affords nonproblematic probability judgments about θ. Viewed from the perspective of the *Stooge*, the *Statistician*'s nonproblematic uncertainty about θ expresses her/his problematic uncertainty about π. But how does the *Statistician* understand the random variable θ so that, without violating the *Total Evidence* requirement, her/his uncertainty about π is related to her/his uncertainty about θ? *Total Evidence* obliges the rational agent to formulate personal probabilities relative to a space of possibility consistent with *all* her/his evidence.

(2) Adapt the requirements for "what you know" to a less than logically omniscient agent. One way to do this is to change the *closure* conditions for what probabilistic assessments rationality demands of a coherent agent. Hacking[4] signals this idea; Garber[5] and Gaifman[6] provide variants of this strategy, as does de Finetti[7] with his theory of *coherence*, which we illustrate below in section II. Then, consonant with Savage's challenge, the agent's uncertainty about the digits of π is no different in kind than the agent's uncertainty about the

[3] I. J. Good, "Twenty-seven Principles of Rationality (#679)" (1971), in *Good Thinking: The Foundations of Probability and Its Applications* (Minneapolis: Minnesota UP, 1983), pp. 15–19.

[4] Hacking, *op. cit.*

[5] Daniel Garber, "Old Evidence and Logical Omniscience in Bayesian Confirmation Theory," in John Earman, ed., *Testing Scientific Theories* (Minneapolis: Minnesota UP, 1983), pp. 99–131.

[6] Haim Gaifman, "Reasoning with Limited Resources and Assigning Probabilities to Arithmetical Statements," *Synthese*, CXL, 1/2 (May 2004): 97–119.

[7] Bruno de Finetti, *Theory of Probability*, vol. 1 (Chichester, UK: Wiley, 1974).

digits of any other mathematical constant. But how to formalize the concept of possibility for such a boundedly rational agent? What is the normative theory of probability for an agent with bounded rationality?

(3) Modify de Finetti's criterion of coherence, which is a dichotomous distinction between coherent and incoherent judgments of personal probability, to accommodate *degrees of incoherence*.[8] Thus, as Savage's comment suggests, the agent's judgment about the digits of π is represented by an incoherent probability assessment. But the modified theory allows for reasoning with incoherent judgments and provides the agent with guidance how to use, for example, ordinary calculations to reduce her/his degree of incoherent uncertainty about the digits of π.

Contemporary probability theory, in particular the mathematical theory of personal probability, relies on a mathematical device, a *measure space* $<\Omega, \mathcal{B}, \mathbf{P}>$, which embeds mathematical and logical structural assumptions. We begin our discussion of these three strategies for addressing Savage's challenge by relating them to the three components of a measure space. Following de Finetti's convention, hereafter, we refer to the reasonable person whose uncertainty about mathematical and logical propositions is the subject of Savage's challenge with the pronoun, "YOU."

The first component of a measure space, $\Omega = \{\omega_i\colon i \in \mathrm{I}\}$ is a partition of YOUR space of serious possibilities, indexed by a set I. The ω_i are called *states*. This attribution as so-called "states" does not require special metaphysical features for the elements ω_i of the partition. These states need not be *atomic* in an absolute sense. Upon further reflection of YOUR opinions, YOU might refine the space, for example, by using a finer partition $\Omega' = \{\omega'_j\colon j \in \mathrm{J}\}$ where each $\omega_i \subseteq \Omega'$. YOU might need to refine Ω when considering, for example, a new random quantity that is not defined with respect to Ω. With respect to Savage's challenge, the problems for YOU in formulating Ω include, for example, that you are unsure whether YOU have succeeded identifying a partition: YOU are unsure whether different elements of Ω are disjoint and whether their union exhausts all the possibilities YOU judge are serious.

$\mathcal{B}$ is a Boolean (sigma) field of subsets of Ω. The elements of $\mathcal{B}$ are the abstract *events* over which YOUR uncertainty is to be represented with a probability function. As we illustrate, below in

[8] Mark J. Schervish, Teddy Seidenfeld, and Joseph B. Kadane, "Measures of Incoherence: How Not to Gamble If You Must, with Discussion," In J. M. Bernardo et al., eds., *Bayesian Statistics* 7 (New York: Oxford, 2003), pp. 385–401.

section II, strategy (2) for responding to Savage's challenge is to relax the conditions that $\mathcal{B}$ is as large as a field of sets. That creates some elbow room for having uncertainty about what is otherwise incorporated as part of the mathematical background assumptions of a measure space.

P is a (countably additive) probability over $\mathcal{B}$ used to represent YOUR uncertainty. We express Savage's challenge to YOU in representing your uncertainty about logical/mathematical constants as follows. In addition to the events that constitute the elements of $\mathcal{B}$, the received theory of mathematical probability introduces a class χ of (possibly bounded) random variables X as ($\mathcal{B}$-measurable) real-valued functions from Ω to $\Re$. Denote by $E_{\mathbf{P}}[X]$ the **P**-expected value of the random variable X. Let $\mathbf{I}_G$ be an indicator function for an *event* G. That is,

$$\mathbf{I}_G(\omega) = 1 \text{ if } \omega \in G \text{ and } \mathbf{I}_G(\omega) = 0 \text{ if } \omega \in G^c.$$

Then $E_{\mathbf{P}}[\mathbf{I}_G] = \mathbf{P}(G)$. Thus, in the received theory, probability is an instance of mathematical expectation. But in the received theory of personal probability, π is a constant variable. It takes the same value in each *state*: $\pi(\omega) = \pi$. So, $E_{\mathbf{P}}[\pi] = \pi$. YOU are required to know π. However, under strategy (3) (as explained in section III), in response to Savage's challenge YOU use an *incoherent* expectation function in order to model YOUR uncertainty about mathematical propositions.

Reflect on Savage's challenge in some detail. Let $X\pi_6$ be the variable whose value is the sixth decimal digit of π. Here, we emphasize the point that YOUR uncertainty about the decimal representation of π may occur without having to consider a "remote" digit. In an ordinary measure space $X\pi_6$ is the constant 2, independent of ω, because π is a constant whose value is independent of the elements of Ω. In an ordinary measure space, with probability 1 the event "$X\pi_6 = 2$" obtains, since as a mathematical result, it obtains in each state ω. Thus, in any ordinary measure space, there is no elbow room for a nonextreme probability about $X\pi_6$ or an expectation other than 2 for its value. Savage's admonition applies:

> For the postulates of the theory imply that you should behave in accordance with the logical implication of all that you know.

The construction of an ordinary measure space requires that you know what constant π is. That fact is part of the mathematical knowledge taken as background also in order to formulate probability *values* in a measure space, as we illustrate, next.

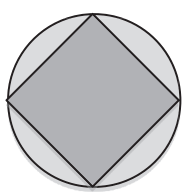
Example 1. Here is an illustration of the use of the mathematical background knowledge for a measure space for giving probability *values*. Consider a problem in probability that relies on three familiar bits of knowledge from high-school geometry.

The area of a circle with radius r equals πr^2.
The area of a square is the square of the length of its side.
The Pythagorean Theorem: Given a right triangle, with side lengths a and b and hypotenuse length c, then $a^2 + b^2 = c^2$.

Let Ω be the set of points interior to a circle **C** with radius r. A point from Ω is chosen at random, with a uniform probability: equal probability for congruent subsets of **C**. Let $\mathcal{B}$ be the algebra of geometric subsets of **C** generated by ruler-and-compass constructions. That is, YOUR personal probability **P** is uniform over these geometric subsets Ω: congruent regions that belong to $\mathcal{B}$ have equal probability. YOU understand that YOUR probability that the random point is contained in a region **S** (for a region **S** that is an element of $\mathcal{B}$) is the ratio of the *area*(**S**) to the *area*(**C**). YOU are aware that YOUR probability of the event "The random point is in **S**" is the fraction *area*(**S**)$/\pi r^2$.

Let **S** be a square inscribed inside the circle **C**. (See Figure 1.) Then by the Pythagorean Theorem and the rule for the area of a square, area(**S**) $= 2r^2$. So, YOU are aware that YOUR probability that the random point is in the square **S** is $2/\pi$. Suppose YOU are aware that the first five decimal digits in the expansion of π are 3.14159. But YOU cannot identify the sixth decimal digit of π. Using the familiar long-division algorithm, then you are unable to calculate precisely YOUR personal probability $(2/\pi)$ beyond the first four digits (0.6366) that the random point is in **S**. YOU know that the fifth digit is either 1 or 2. But, for instance, then YOU are unable to answer whether a bet on the random point is in **S** at odds of .63662:.36338 is favorable, fair, or unfavorable for YOU.$\lozenge_{\text{Example}}$

Figure 1



Thus, the challenge Savage poses affects both *the numerical values* that YOU can identify for YOUR (coherent) probability assessments, as well as the *random quantities* to which YOU can assign a coherent probability assessment.[9] With strategy (1), next we illustrate how to convert this "bug" into a "feature" that opens the door to using commonplace numerical methods as a response to Savage's challenge.

I. STRATEGY (1)

We extend Example 1 to illustrate strategy (1): Loosen the grip of the *Total Evidence Principle.* Use a *Statistician's Stooge* to replace the original uncertain quantity $X\pi_6$ with a different one, θ, that the *Stooge* knows (but YOU do not know) is coextensive with $X\pi_6$. Then YOU may hold nonextreme but coherent probabilities about the substitute variable θ. In this way, familiar numerical methods, including Monte Carlo methods, permit YOU to learn about $X\pi_6$ by shifting the failure of the Total Evidence principle to the *Stooge.*

Example 1 (continued). As an instance of I. J. Good's *Statistician's Stooge*, YOUR assistant, the *Stooge,* creates an elementary statistical estimation problem for the quantity $2/\pi$ using *iid* repeated draws from the uniform distribution on a circle **C**. The *Stooge* chooses **C** to be the circle with center at the origin (0, 0) and radius $r = \sqrt{2}$. Then the inscribed square **S** has corners with coordinates (± 1, ± 1). Let $X_i = (X_{i_1}, X_{i_2})$ ($i = 1, \ldots, n$) be n random points drawn by the *Stooge* using the uniform distribution on **C**. After each draw the Stooge determines whether or not $X_i \in \mathbf{S}$, that is, whether or not both inequalities obtain: $-1 \leq X_{i_j} \leq +1$ ($j = 1, 2$), which involves examining only the first significant digit of X_{i_j}.

Now, the *Stooge* tells YOU whether event Y occurs on the i^{th} trial, $Y_i = 1$, if and only if $X_i \in \mathbf{S}$ for a region **S**. But all the *Stooge* tells

[9] Example 1 opens the door also to a discussion of *higher-order* probabilities. YOU might try to assign a second-order personal probability distribution $\mathbf{P}^*$ to the quantity $2/\pi$ in order to represent the added higher-order uncertainty you have in YOUR first-order uncertainty **P** that the random point is in the region **S**. Higher-order probability is a topic beyond the focus of this essay. Here, we express our agreement with the Savage-Woodbury rejoinder—Savage, *The Foundations of Statistics*, 2nd ed. (New York: Dover, 1954/1972), p. 58. That rejoinder questions whether such a higher-order personal probability has operational content. The Savage-Woodbury response establishes that $\mathbf{P}^*$ provides YOU with a resolution of your first-order uncertainty: Use $\mathbf{P}^*$ to create an expected value for $2/\pi$, just as you would use personal probability to determine an expected value for any random quantity. Then this expected value is your first-order expected value for $2/\pi$, and YOU have no added uncertainty about your (first-order) probability that the random point is in **S**. Then there is no residual higher-order uncertainty.

Example 1 also opens the door to upper and lower previsions that govern one-sided gambles. We consider this in connection with de Finetti's *Fundamental Theorem of Previsions,* which we discuss in section II in connection with Example 2. We indicate why upper and lower previsions do not resolve Savage's challenge, either.

YOU about the region **S** is that it belongs to the algebra $\mathcal{B}$. Then the Y_i form an *iid* sequence of Bernoulli(θ) variables, where θ is the area(**S**)/2π. As it happens, $\theta = 2/\pi$. But this identity is suppressed in the following analysis, with which both YOU and the *Stooge* concur.

YOU and the *Stooge* know that $\sum_{i=1}^{n} Y_i$ is Binomial(n,θ). Let $\overline{Y}_n = \sum_{i=1}^{n} Y_i/n$ denote the sample average of the Y_i. $\overline{Y}_n$ is a *sufficient statistic* for θ, that is, a summary of the n draws X_i that preserves all the relevant evidence in a coherent inference about θ based on the data of the n-many *iid* Bernoulli(θ) draws.

The *Stooge* samples with $n = 10^{16}$, obtains $\overline{Y}_n = 0.63661977236$, and carries out ordinary Bayesian reasoning with YOU about the Binomial parameter θ using YOUR "prior" for θ. According to what the *Stooge* tells YOU, θ is an uncertain Bernoulli quantity of no special origins. YOU tell the *Stooge* your "prior" opinion about θ. For convenience, suppose that YOU use a uniform conjugate Beta(1, 1) "prior" distribution for θ, denoted here as $\mathbf{P}(\theta)$. So, the *Stooge* reports, given these data, YOUR "posterior" probability is greater than .999, that $0.63661971 \leq \theta \leq 0.63661990$. Then, since the *Stooge* knows that $\theta = 2/\pi$, the *Stooge* reports for YOU that the probability is at least .999 that the sixth digit of π is 2. Of course, in order for YOU to reach this conclusion you have to suppress the information that **S** is an inscribed square within **C**, rather than some arbitrary geometric region within the algebra of ruler-and-compass constructions. The *Stooge* needs this particular information, of course, in order to determine the value of each Y_i.$\lozenge_{\text{Example}}$

This technique, strategy (1), generalizes to include the use of many familiar numerical methods as a response to Savage's question: How do YOU express uncertainty about a mathematical term τ? The numerical method provides evidence in the form of a random variable, Y, whose value $Y = y$ is determined by an experiment with a well-defined likelihood function, $\mathbf{P}(Y{=}y \mid \theta)$, that depends upon a parameter θ, known to the *Stooge* but not to YOU to equal the problematic quantity τ. YOU express a coherent "prior" probability for θ, $\mathbf{P}(\theta)$. By Bayes's Theorem, YOUR "posterior" probability, $\mathbf{P}(\theta \mid Y{=}y)$ is proportional to the product of this likelihood and "prior":

$$\mathbf{P}(\theta \mid Y{=}y) \propto \mathbf{P}(Y{=}y \mid \theta)\, \mathbf{P}(\theta).$$

As Good notes, playing fast and loose with the Total Evidence principle—in the example, by permitting the *Stooge* to suppress the problematic information that $\theta = \tau$—allows YOU, a coherent Bayesian statistician, to duplicate some otherwise non-Bayesian, Classical statistical inferences. For instance, a Classical α-level Confidence Interval for a quantity θ based on a random variable X, $\mathrm{CI}^{\alpha}_{lower}(X) \leq \theta \leq \mathrm{CI}^{\alpha}_{upper}(X)$ becomes a Bayesian posterior probability α for the same interval estimate of θ given X, by suppressing the observed value of the random variable, $X = x$, and leaving to the *Stooge* the responsibility of filling in that detail.

What troubles us about this approach as a response to Savage's challenge is that YOUR coherent uncertainty about the substitute parameter θ may reflect very little of what YOU know about the problematic quantity τ. Employing the *Stooge*, as above, allows YOU to express a coherent prior probability for θ. In Example 1, θ is the ratio: area of some arbitrary rule-and-compass region chosen (by the *Stooge*) from $\mathcal{B}$ divided by 2π, $\tau = 2/\pi$, and, unknown to YOU, $\theta = \tau$. But YOUR prior for θ, when that quantity is identified to YOU as just some region chosen by the *Stooge*, may have very little in common with YOUR uncertainty about τ, which depends upon the problematic information that **S** is the inscribed square and which the *Stooge* conveniently suppresses for YOU.

II. STRATEGY (2)

In this section we examine an instance of strategy (2)—modify the closure conditions on the space of uncertain events in order to avoid requiring YOU are logically/mathematically omniscient. Hacking (1967) responds to Savage's challenge this way. Here, we review de Finetti's (1974) theory of *coherent Previsions*: $P(\cdot)$ as an instance of this strategy.

In de Finetti's theory, YOU are required to offer a *fair price*, a *prevision* $P(X)$, for buying and selling the random variable X. X is defined for/by YOU with respect to a partition Ω. That is, for each state $\omega \in \Omega$, $X(\omega)$ is a well-defined real number. That is, the function $X{:}\Omega \rightarrow \Re$ is known to YOU. (In connection with the *Dutch Book* argument, de Finetti often refers to YOU as the *Bookie.*) To say that $P(X)$ is YOUR fair price for the random quantity X means that YOU are willing to accept all contracts of the form $\beta_{X,P(X)}[X - P(X)]$, where an opponent (called the *Gambler*) chooses a real-value, $\beta_{X,P(X)}$. This term, $\beta_{X,P(X)}$, is constrained in magnitude in order to conform to YOUR wealth, but allowed to depend on both the variable X and YOUR prevision for X. With $|\beta| > 0$ small enough to fit YOUR budget, YOU are willing to engage in the following contracts:

when $\beta > 0$ YOU agree to pay $\beta P(X)$ in order *to buy* (that is, to receive) βX in return;

when $\beta < 0$ YOU agree to accept $\beta P(X)$ in order *to sell* (that is, to pay) βX in return.

For finitely many contracts YOUR outcome is the sum of the separate contracts.

$$\sum_{i=1}^{n}\beta_i[X_i(\omega) - P(X_i)].$$

In de Finetti's theory, the state space $\Omega = \{\omega\}$ is formed by taking all the mathematical combinations of those random variables $\chi = \{X\}$ that YOU have assessed with YOUR *previsions*. We illustrate this technique in Example 2, below.

Definition. YOUR *Previsions* are collectively *incoherent* provided that there is a finite combination of acceptable contracts with uniformly negative outcome—if there exists a finite set $\{\beta_i\}$ ($i = 1, \ldots, n$) and $\varepsilon > 0$ such that, for each $\omega \in \Omega$,

$$\sum_{i=1}^{n}\beta_i[X_i(\omega) - P(X_i)] < -\varepsilon.$$

With this choice of $\{\beta_i\}$ the *Gambler* has created a sure loss for YOU—a *Dutch Book*. Otherwise, if no such combination $\{\beta_i\}$ exists, YOUR previsions are *coherent*.

Let $\chi = \{X_j: j \in J\}$ be an arbitrary set of variables, defined on Ω. What are the requirements that coherence imposes on YOU for giving coherent previsions to each random quantity in the set χ? That is, suppose YOU provide previsions for each of the variables X in a set χ where each variable X is defined with respect to Ω, that is, the function $X: \Omega \rightarrow \Re$ is well defined for each X. When are these a coherent set of previsions?

De Finetti's *Theorem of Coherent Previsions*:[10]

YOUR *Previsions* are coherent *if and only if* there is a (finitely additive) probability $\mathbf{P}(\cdot)$ on Ω with YOUR *Previsions* equal to their $\mathbf{P}$-expected values.

$$P(X) = \mathrm{E}_{\mathbf{P}}[X].$$

This theorem yields the familiar result that, when all the variables in χ are indicator functions—when all of the initial gambles are simple bets on events—YOUR previsions are immune to the Gambler having a strategy for making a Book against you if and only if your previsions are a (finitely additive) probability.

[10] de Finetti, *Probabilismo: Saggio critico sulla teoria della probabilità e sul valore della scienza* (Naples, Italy: Perrella, 1931), translated as "Probabilism: A Critical Essay on the Theory of Probability and on the Value of Science," *Erkenntnis*, XXI, 2/3 (September 1989): 169–223; and de Finetti, "La prévision: ses lois logiques, ses sources subjectives," *Annales de L'Institut Henri Poincaré*, VII (1937): 1–68, translated as (and with new notes by the author) "Foresight: Its Logical Laws, Its Subjective Sources," in Henry E. Kyburg, Jr., and Howard E. Smokler, eds., *Studies in Subjective Probability*, 2nd ed. (Huntington, NY: Krieger, 1980).

De Finetti's theory of coherent previsions commits YOU to having precise previsions for all variables in the linear span—for all linear combinations—of those variables $X \in \chi$ that you have already assessed with previsions. As we explain, below, this is a different closure condition than requiring YOU to have previsions determined even for all events in the smallest logic/algebra generated by Ω.

Again, suppose YOU provide coherent previsions for all variables X in the set χ. Let Y be another variable defined with respect to Ω but not necessarily in χ.

Let: $\underline{A} = \{X\colon X(\omega) \leq Y(\omega)$ and X is in the *linear span* of $\chi\}$
$\overline{A} = \{X\colon X(\omega) \geq Y(\omega)$ and X is in the *linear span* of $\chi\}$

Let: $\underline{P}(Y) = sup_{X \in \underline{A}}\, P(X)$ and $\overline{P}(Y) = inf_{X \in \overline{A}}\, P(X)$.

De Finetti's *Fundamental Theorem of Previsions*:

Extending YOUR previsions P to P^* in order to give a coherent prevision for Y, $P^*(Y)$, allows it to be any (finite) number from $\underline{P}(Y)$ to $\overline{P}(Y)$.

Outside this interval, the extension P^* is incoherent.

Next we illustrate these two results of de Finetti and explain their relevance to Savage's challenge.

Example 2. Consider a roll of a six-sided die with faces numbered in the usual way, 1, 2, 3, 4, 5, 6, and with opposite sides always summing to 7. Suppose YOU think about the following four events (which define the set χ) and identify YOUR *Previsions* in accord with the assessment that the die is *fair*:

$$P(\{1\}) = 1/6;\ P(\{3,6\}) = 1/3;\ P(\{1,2,3\}) = P(\{1,2,4\}) = \tfrac{1}{2}.$$

The set of events for which YOUR coherent prevision is already determined by the previsions for these four events is given by the *Fundamental Theorem*. That set does *not* form an algebra. Only 24 of 64 events (only 12 pairs of complementary events) have determinate previsions.

For instance, by the *Fundamental Theorem*:

$\underline{P}(\{6\}) = 0 < \overline{P}(\{6\}) = 1/3;$
likewise $\underline{P}(\{4\}) = 0 < \overline{P}(\{4\}) = 1/3;$
however, $P(\{4,6\}) = 1/3.$

The smallest algebra for the four events in χ is the power set of all 64 subsets of Ω. Thus, de Finetti's theory of *coherence* does *not* require that YOUR previsions are well defined for all the propositions in the elementary logic formed from YOUR beliefs about the constituents. YOU do not have to close the set of YOUR previsions under even sentential logical operations. For instance, YOU are not required to provide a well-defined prevision for an event that is the intersection of two events each of which you have assessed with well-defined previsions. In Example 2, YOU give determinate previsions $P(\{3,6\})$ and $P(\{1,2,3\})$, but are not required by coherence to assess $P(\{3\})$. Alas, however, this approach through de Finetti's *Fundamental Theorem* does not solve YOUR question of how to depict uncertainty about mathematical/logical constants.$\Diamond_{\text{Example}}$

Example 3. For convenience, label the four events in χ: $F_1 = \{1\}$, $F_2 = \{3,6\}$, $F_3 = \{1,2,3\}$, and $F_4 = \{1,2,4\}$. Consider the following specific sentential proposition, H, about which we presume YOU are unsure of its validity—until, that is, you calculate truth tables.

$$H: \qquad [(F_2 \vee [F_1 \wedge (F_4 \vee F_3)]) \rightarrow [(F_2 \vee F_1) \wedge (F_2 \vee F_4)]]$$

Analogous to the variable $X\pi_6$, the sixth decimal digit in π, the indicator variable $\mathbf{I}_H$ is a constant: it takes the value 1 for each state in Ω. $1 = \mathbf{I}_\Omega \leq \mathbf{I}_H$. So, by the *Fundamental Theorem,* in order to be coherent YOUR prevision must satisfy $P(\mathbf{I}_H) = 1$. Assume that, prior to a truth-table calculation, YOU are unsure about H. Alas, de Finetti's theory of coherent previsions leaves YOU no room to express this uncertainty. The closure of coherent previsions required by the *linear span* of the random variables that YOU have coherently assessed does *not* match the *psychological* closure of your reasoning process.

Here is the same problem viewed from another perspective.

Example 3 (continued). Garber (1983) suggests YOU consider the sentential form of the problematic hypothesis as a way of relaxing the structural requirements of logical omniscience.

$$H: \qquad [(F_2 \vee [F_1 \wedge (F_4 \vee F_3)]) \rightarrow [(F_2 \vee F_1) \wedge (F_2 \vee F_4)]]$$

This produces the schema:

$$\mathcal{H}': \qquad [\mathcal{P} \vee (\mathcal{Q} \wedge (\mathcal{R} \vee \mathcal{S}))] \rightarrow [(\mathcal{P} \vee \mathcal{Q}) \wedge (\mathcal{P} \vee \mathcal{R})]$$

Evidently $\mathcal{H}'$ is neither a tautology nor a contradiction. So, each value $0 \leq P(\mathcal{H}) \leq 1$ is a coherent prevision, provided that we have the full set of truth-value interpretations for the sentential variables $\mathcal{P}$, $\mathcal{Q}$, $\mathcal{R}$, and $\mathcal{S}$.$\Diamond_{\text{Example}}$

The replacement of H by $\mathcal{H}$ ignores the underlying mathematical relations among the variables in H. Suppose that YOU assess YOUR prevision for $\mathcal{H}$, $P(\mathcal{H}) = .6$. Does YOUR psychological state of uncertainty about H match the requirements that *coherence* places on a prevision, $P(\mathcal{H}) = .6$? Do *YOU* identify $\mathcal{H}'$ as the correct variable for what you are thinking about when you are reflecting on YOUR uncertainty about H, before you do the calculations that reveal H is a logical constant? We think the answer is "No."

The same problem recurs when, instead of imposing the norms merely of a sentential logic, as in Garber's suggestion, we follow Gaifman's (2004) intriguing proposal for reasoning with limited resources. Gaifman offers YOU a (possibly finite) collection P of *sentences* over which you express your degrees of belief. As Gaifman indicates, in his approach sentences are the formal stand-ins for Fregean *thoughts*—"senses of sentences," as he puts it (2004, p. 102). This allows YOU to hold different degrees of uncertainty about two *thoughts* provided that they have different *senses*. In Gaifman's program, YOUR opinions about sentences in P are governed by a restricted logic. He allows for a *local algebra* of sentences that are provably equivalent in a restricted logic. Then YOUR assessments for the elements of P might not respect logical equivalence, as needed in order to escape the clutches of logical omniscience. Just as with de Finetti's rule of closure under the linear span of assessed events, also in Gaifman's system of a *local algebra* YOU are not required to assess arbitrary well-formed subformulas of those in P.

We are unsure just how Gaifman's approach responds to Savage's challenge. First, as a practical matter, we do not understand what YOUR previsions for such sentences entail when previsions are used as betting rates. When YOU bet on a sentence $\boldsymbol{s}$ (in a *local algebra*), what are the payoffs associated with such a bet? That is, how does a *local algebra* fix the payoffs when YOU bet on $\boldsymbol{s}$ with prevision $P(\boldsymbol{s})$? It cannot be that the truth conditions for $\boldsymbol{s}$ determine the payoffs for the bet. That way requires YOU to be logically omniscient if you are coherent, of course.

Second, and more to the point of Savage's challenge, we do not see why YOUR uncertainty about mathematical propositions should match the normative constraints of an algebra closed under some finite number of iterations of a given rule of inference. Why should YOUR uncertainty over mathematical propositions match what is provable in a restricted *local algebra* of the kind sketched by Gaifman? Might it not be that YOU recall the seventh digit of π but not the sixth? Then, mimicking YOUR uncertainty about the digits of π with a restricted deductive system that generates the digits of π in a proof, according to a computation of π, will not capture YOUR uncertainty about π.

In Levi's terms,[11] YOUR *commitments* to having coherent (precise) previsions according to de Finetti's norms of coherence do not match YOUR *performance* when assessing YOUR uncertainty about

[11] Isaac Levi, *The Fixation of Belief and Its Undoing: Changing Beliefs Through Inquiry* (New York: Cambridge, 1991).

mathematical propositions. Nor does YOUR performance match the norms of a sentential logic, as per Garber's proposal. Nor does YOUR performance match the norms of a *local algebra*, as per Gaifman's proposal. What reason makes plausible the view that YOUR thinking about a mathematical proposition, your actual performance when judging the value of $X\pi_6$, matches the commitments of any such normative theory? We are doubtful of strategy (2)!

III. STRATEGY (3)

One feature common to strategies (1) and (2) is the goal of showing that YOU are coherent when you hold nonextreme personal probabilities for mathematical propositions.

Strategy (1) allows YOU to replace a problematic mathematical variable, for example, $X\pi_6$, one that is constant across the space of all possibilities (Ω), with another random variable θ that is not problematic in the same way. Numerical methods for computing the problematic variable, $X\pi_6$, then can be modeled as ordinary statistical experiments generating data Y about θ. There is no incoherence when YOU use nonextreme personal probabilities for θ, given Y. But, as we saw, in addition to failure to adhere to the Total Evidence principle, the effectiveness of strategy (1) depends upon YOUR willingness to use your prior probability for θ to express your thinking about $X\pi_6$.

Strategy (2) allows YOU to replace the familiar algebra $\mathcal{B}$ of a measure space with some other mathematical structure that can support a different set of coherent personal probability assessments—a set that is less demanding on YOUR logical reasoning abilities. For de Finetti, that other mathematical structure is the linear span formed by those previsions you are willing to make. For Garber it is the structure of a sentential logic. For Gaifman it is a *local algebra.* Though each of these might capture some aspect of YOUR thinking about a mathematical proposition, why should YOU think according to the norms of any one of these alternative mathematical structures? None of them is intended as a realistic psychological theory of how YOUR mind reasons.

We propose, instead, strategy (3): Concede that, regarding uncertainty about mathematical and logical propositions, despite the phenomenological similarities with uncertainty about nonconstant "empirical" variables, nonextreme *previsions* for mathematical propositions are incoherent. This is exactly what Savage points out is the problem with the theory of Personal Probability.

> *That is, we concede that YOUR commitment to being coherent is not discharged by what is an entirely predictable shortfall in YOUR performance.*

However, the problem is exacerbated by the fact that de Finetti's distinction between *coherent* and *incoherent* previsions is dichotomous. Perhaps a more nuanced theory of incoherence can guide incoherent thinkers on how to reason without abandoning their commitment to coherence? That is the core idea for strategy (3).

In several of our papers we develop a theory of *degrees of incoherence.*[12] When the *Gambler* can make a Book against the *Bookie*'s incoherent previsions, then many Books can be made. The different Books may be compared by *scaling* the (minimum) *sure gain* to the Gambler—or equivalently scaling the minimum *sure loss* to the Bookie, or adopting a Neutral index, which incorporates both perspectives.

Here are three indices that may be used to scale the sure gains/losses in a Book. For simplicity, in the following discussion we scale the finite set of gambles in a Book using the sum of the individually scaled gambles. This is a special case of our general theory.

> *Rate of Loss (for the Bookie)*: Scale the minimum sure loss to the *Bookie* by the total amount the *Bookie* is compelled to wager from the *Gambler*'s strategy.

What proportion of the *Bookie*'s budget can the *Gambler* win for sure?

> *Rate of Profit (for the Gambler)*: Scale the sure gain to the *Gambler* by the total amount used in the *Gambler*'s strategy.

What proportion of the *Gambler*'s stake does the *Gambler* have to escrow to win one unit for sure from the *Bookie*?

> *A Neutral Rate*: Scale the sure loss to the *Bookie* by the combined amounts (the total stake) wagered by both players according the *Gambler*'s strategy.

As we explain in our (2003) paper, this index is better designed than either of the first two for assessing incoherent previsions for constants. With the Neutral Rate, if $X_c(\omega) = c$ is a constant variable and $P(X_c)$ is a prevision for X_c, then the degree of incoherence for this one prevision is $|c - P(X_c)|$.

Relative to each of these indices, the *rate of incoherence* for an incoherent *Bookie*'s previsions is the greatest (scaled) loss/gain that the *Gambler* can achieve across the different strategies for making a Book.

Example 4: Illustrating Differences among These Three Rates of Incoherence. Consider a 3-element state space, $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Let $\chi = \{\mathbf{I}_i\colon i = 1, 2, 3\}$ be the set of the three indicator functions for

[12] For an overview, see Schervish, Seidenfeld, and Kadane, *op. cit.*

the three elements of Ω. And let the following be three incoherent prevision functions over χ.

$$P_1(\omega_i) = P_1(\mathbf{I}_i) = <0.5, 0.5, 0.5>, \text{ for } i = 1, 2, 3.$$
$$P_2(\omega_i) = P_2(\mathbf{I}_i) = <0.6, 0.7, 0.2>, \text{ for } i = 1, 2, 3.$$
$$\text{and} \quad P_3(\omega_i) = P_3(\mathbf{I}_i) = <0.6, 0.8, 0.1>, \text{ for } i = 1, 2, 3.$$

With prevision $P_j(\cdot)$, $j = 1, 2$, or 3, each contract of the form $\alpha_i(\mathbf{I}_i - P_j(\omega_i))$ is judged fair. In all three cases ($j = 1, 2, 3$), $P_j(\omega_1) + P_j(\omega_2) + P_j(\omega_3) = 1.5$, and not $= 1$. Clearly, these are incoherent previsions. For the first incoherent prevision, $P_1(\cdot)$, all three rates of incoherence lead the *Gambler* to the same strategy, bet against the incoherent *Bookie*$_1$ with equal stakes on all three states, $\alpha_i = (1, 1, 1)$. For the second incoherent prevision, $P_2(\cdot)$, both the Rate of Loss and the Neutral Rate are maximized with the *Gambler* using the equal-stakes strategy, $\alpha_i = (1, 1, 1)$. But the Rate of Profit against incoherent *Bookie*$_2$ is maximized with the strategy $\alpha_i = (1, 1, 0)$, that is, gamble only on the first two states, and with equal stakes. For the third incoherent prevision, $P_3(\cdot)$, the Rate of Loss is maximized against *Bookie*$_3$ with the *Gambler*'s strategy of equal stakes on all three states, $\alpha_i = (1, 1, 1)$, whereas the other two rates of incoherence are maximized with the other strategy $\alpha_i = (1, 1, 0)$.$\Diamond_{\text{Example}}$

Thus, the three rates lead the *Gambler* to three different combinations of strategies. These are different ways to index a *rate of incoherence*. Of course, each coherent prevision has a 0-rate of incoherence with each index.

We first developed our ideas about rates of incoherence in order to engage familiar debates about Bayesian versus Classical Statistical procedures. Bayesians argue that where a particular Classical procedure is incoherent, therefore it is unacceptable. But this is a coarse-level analysis. We inquire, instead, how incoherent is the Classical procedure. Since Classical statistical procedures are often simple to calculate, they can be warranted in the special case when the rate of incoherence is small and a rival, full Bayesian analysis is computationally infeasible. We provide an illustration of this analysis in our (2000), where we investigate the rate of incoherence of fixed α-level hypothesis tests regardless of sample size.[13]

How do we propose to use our ideas about rates of incoherence to address Savage's challenge of how to use probabilities to formalize uncertainty about mathematical propositions? In the spirit of de Finetti's *Fundamental Theorem*, the following result,

[13] Schervish, Seidenfeld, and Kadane, "A Rate of Incoherence Applied to Fixed-Level Testing," *Philosophy of Science*, LXIX, S3 (September 2000): S248–64.

reported in section 6 of our (2003), explains how to calculate a prevision for a new variable without increasing YOUR existing rate of incoherence.

Assume YOU assess previsions for each element of a (finite) partition $\pi = \{h_1, \ldots, h_m\}$, with values $P(h_i) = p_i$, $i = 1, \ldots, m$. YOU are asked for YOUR prevision $P(Y)$ for a (π-measurable) variable Y, with $Y(h_i) = c_i$.

- Calculate a pseudo-expectation using *YOUR* possibly incoherent previsions over π: $P(Y) = \sum_i p_i c_i$
- Then you will not increase *YOUR Rate of Incoherence* extending your previsions to include the new one for Y, $P(Y) = \sum_i p_i c_i$

When YOU are coherent, YOUR rate of incoherence is 0. Then pseudo-expectations are expectations, and the only way to extend YOUR previsions for a new variable, while preserving YOUR current 0-rate of incoherence, is to use the pseudo-expectation algorithm. However, when YOU are incoherent, there are other options for assessing $P(Y)$ without increasing YOUR rate of incoherence. But, without knowing how incoherent YOU are, still YOU can safely use the pseudo-expectation algorithm and be assured that your rate of incoherence does not increase. The pseudo-expectation algorithm is *robust*!

One intriguing case of this result arises when Y is the variable corresponding to a *called-off* (conditional) gamble.[14] Then using a pseudo-expectation with respect to YOUR (possibly) incoherent previsions for Y suggests how to extend the principle of *confirmational conditionalization*[15] to include incoherent conditional previsions. When YOU hypothesize expanding your *corpus of knowledge* to include the new evidence ($X = x$), YOUR possibly incoherent previsions $P(\cdot)$ become $P(\cdot \mid X = x)$, as calculated according to the Bayes algorithm for pseudo-expectations.

This leads to the following Corollary, which is an elementary generalization of familiar results about the asymptotic behavior of a coherent posterior probability function given a sequence of identically, independently distributed (*iid*) variables.[16]

[14] We discuss this in section 6 of Schervish, Seidenfeld, and Kadane, "Two Measures of Incoherence," Technical Report #660, Department of Statistics, Carnegie Mellon University (1997).

[15] See Levi, *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance* (Cambridge: MIT, 1980).

[16] See Savage, *Foundations of Statistics*, p. 141, Theorem 1, for the special case of a finite parameter space, and Doob's theorem, as reported by Schervish, *Theory of Statistics* (New York, Springer-Verlag, 1995), T.7.78, p. 429, for the general version, as used here.

Corollary. Let Θ be a finite-dimensional parameter space. Consider a nonextreme, *pseudo-prior density function* $p(\theta) > 0$, which may be incoherent. Suppose, however, that a *pseudo-likelihood density function* $p(X = x \mid \theta)$ has a 0-rate of incoherence; that is, these conditional probabilities are coherent. Suppose, also, they are different conditional probability functions for different values of θ. Let $X_i (i = 1, \ldots)$ form a sequence of conditionally *iid* variables, given θ, according to $p(X = x \mid \theta)$. Use the *pseudo-Bayes-algorithm* to create a sequence of *pseudo-posterior functions* $p_n(\theta \mid X_1, \ldots, X_n)$, $n = 1, \ldots$.

Then, almost surely with respect to the true state, $\theta^* \in \Theta$, the Neutral rate of incoherence for the *pseudo-posterior* converges to 0, and that *pseudo-posterior* concentrates on θ^*.

Example 1 (concluded). Reconsider the version of Example 1 involving *iid* repeated sampling of the bivariate variable X, a point randomly chosen from a circle **C**. **S** is a particular inscribed square. Let $Y_i = 1$, if $X_i \in \mathbf{S}$, and $Y_i = 0$, if $X_i \notin \mathbf{S}$. Let $\theta = 2/\pi = P(Y{=}1 \mid \theta)$. Suppose YOU assign a smooth but incoherent pseudo-prior to θ, for example, use a Beta(1, 1) pseudo-prior. Then, given the sequence $Y_n (n = 1, \ldots)$, by the Corollary, the sequence of YOUR pseudo-posteriors, $P_n(\Theta \mid Y_1, \ldots, Y_n)$ converges (even uniformly) to $2/\pi$. With the Neutral Rate, if $X_c(\omega) = c$ is a constant variable and $P(X_c)$ is a prevision for X_c, then the degree of incoherence for this one prevision is $|c - P(X_c)|$. Therefore, almost surely, also the Neutral Rate of incoherence in YOUR pseudo-posterior converges to $0.\Diamond_{\text{Example}}$

Thus, we see how to use data from familiar numerical methods, methods that have well-defined, coherent likelihood functions—as in the growing family of MCMC algorithms—to improve the rate of incoherence in our previsions for mathematical propositions.

IV. SUMMARY

We have reviewed three strategies for addressing the question whether probability theory can be used to formalize personal uncertainty about ordinary mathematical propositions. We posed the problem in the following form. Variable X_c is a mathematical/logical constant that YOU are unable to identify. So, according to the theory of Personal Probability YOUR nonextreme prevision for X_c is incoherent.

(1) Relax the Total Evidence requirement, for example, use I. J. Good's *Statistician's Stooge*, in order to substitute a related variable θ, about which ordinary statistical inference is coherent, for the problematic variable X_c. With the *Stooge*'s help in censuring some empirical information (for example, $X_c = \theta$), you can reason coherently about θ. It is the *Stooge* who converts those conclusions into incoherent

previsions about X_c. But how to match θ against what we are thinking about X_c? What exactly is our *Stooge* reporting to us about θ?

(2) Relax the structure of a measure space in order to accommodate a more psychologically congenial closure condition on the set of variables to be assessed (Hacking, 1967). What fits the bill? De Finetti's use of the linear span in place of an algebra of events does not work. Nor does either Garber's proposal to use sentential logic, or Gaifman's *local algebra*. We do not see how to match YOUR coherent assessments, where you are aware of these, with a domain of propositions defined by mathematical operations. The mathematical operations used for "closing" the domain of propositions form a Procrustean bed against the domain of YOUR coherent assessments.

(3) Concede that nonextreme probabilities for mathematical propositions are incoherent. Then provide normative criteria for reasoning with incoherent previsions in order to show how to reduce YOUR rate of incoherence. The dichotomy between coherent/incoherent assessments appears too coarse to explain how we use, for example, numerical methods to improve our thinking about mathematical quantities. With our approach to Savage's challenge, using the machinery of rates of incoherence, we expand an old Pragmatist idea—one that runs from Peirce through Dewey. We illustrate how to make the operation of a numerical calculation into an experiment whose outcome may be analyzed using familiar principles of statistical inference. Here, we have taken a few, tentative steps in this direction.

We do not know, however, how far our approach goes in addressing the scope of Savage's challenge. For example, a commonplace decision for a mathematician unsure about a specific mathematical conjecture is how to apportion her/his efforts between searching for a proof of the conjecture and searching for a counterexample to the same conjecture. In sections 14.14–14.15 of his (1970), dealing with search problems, DeGroot establishes Bayesian algorithms for optimizing sequential search.[17] Can these algorithms be adapted to the mathematician's decision problem by allowing for some incoherence in her/his assessments about the conjecture? Are the algorithms DeGroot proves optimal also with pseudo-expectations? We take this as a worthy conjecture.

TEDDY SEIDENFELD
MARK J. SCHERVISH
JOSEPH B. KADANE

Carnegie Mellon University

[17] Morris H. DeGroot, *Optimal Statistical Decisions* (New York: McGraw-Hill, 1970).

ERNEST NAGEL AND REDUCTION*

Ernest Nagel first presented his account of theory reduction in science in a chapter in a 1949 philosophy collection.[1] That essay grew into a longer chapter in his 1961 book, *The Structure of Science*, supplemented there with a discussion of emergence and an in-depth analysis of the relations of wholes and parts.[2] Finally, in 1979, Nagel republished a 1970 essay in his collection *Teleology Revisited*, responding to many of the criticisms voiced in the 1960s and 1970s, especially that of Feyerabend, and also touching on some issues in the autonomy of biology.[3] The 1949 essay is of interest as the initial formulation of what continued through 1979 to be a largely unchanging and immensely influential analysis of reduction. The *locus classicus* of Nagelian reduction, however, remains his 1961 chapter, though it is usefully supplemented with some remarks in the 1979 chapter.

In the next few pages I summarize the Nagel reduction analysis, and then in the following two sections comment on a variety of the diverse responses and extensions of Nagel's account, roughly dividing these into the first 40 years of publications and those published in this century. Threaded into this account is a specific thesis that ultimately all attempts at theory reductions in science are "incomplete," "partial," or "patchwork" in character. The nature and degree of the incompleteness varies with the type of science, however, and in physics, because of the "Euclidean" form of its theories, virtually "systematic" or "sweeping" Nagelian-type reductions seem possible. It was the prospect of these sweeping reductions that motivated the original Nagel model of reduction and its application to thermodynamics and statistical mechanics (hereafter, the "canonical SM example"). However, more than sixty years of explorations and refinements of the model strongly support a more "creeping" form of reductions,

* My thanks to James Bogen, Jeremy Butterfield, Stephan Hartmann, John Norton, Sahotra Sarkar, Mark Wilson, Jim Woodward, and the editors of this JOURNAL for constructive comments on various versions of this paper, as well as to the National Science Foundation for support of my research on reduction in the sciences.

[1] Ernest Nagel, "The Meaning of Reduction in the Natural Sciences," in Robert C. Stauffer, ed., *Science and Civilization* (Madison: Wisconsin UP, 1949), pp. 97–135.

[2] Nagel, *The Structure of Science: Problems in the Logic of Scientific Explanation* (New York: Harcourt, Brace and World, 1961).

[3] Nagel, "Issues in the Logic of Reductive Explanations," in *Teleology Revisited and Other Essays in the Philosophy and History of Science* (New York: Columbia, 1979), pp. 95–117.

0022-362X/12/0908-09/534–65

particularly in the biological sciences and the neurosciences. This view, however, does not impugn the importance of such partial reductions. Such partial reductions amount to potentially Nobel Prize-winning accomplishments in unifying and deepening significant areas of scientific investigation. Even in physics, though reductive expositions can give the appearance of a complete systematic reduction, closer inspection reveals that this systematicity fails at the margins, and even where successful the reductions often involve modifications of *both* reduced and reducing theories as various incomplete aspects of the reduction are addressed.

I add further support for the nuances of this thesis by considering a thus-far largely ignored example of a successful and virtually systematic theory reduction, one mentioned several times as being such by Nagel but rarely pursued in the literature. That example is the relation between physical optics and electromagnetic theory, an example which also clarifies several controversial aspects of reductions and indirectly illuminates the canonical SM example first introduced by Nagel. A fuller exposition of the optics example than can be included in the present paper will appear in a separate companion paper.[4] Finally, I conclude with an analysis of what I take to be the main lessons from the extensive discussion of the Nagel model, in the context of a model of partial reductions.

I. NAGEL'S ANALYSIS OF THEORY REDUCTION IN SCIENCE

The initial motivation behind Nagel's analysis of reduction was the success of nineteenth-century mechanics in absorbing and thus reducing other branches of science—whence, perhaps, Nagel's choice of a reduction of classical thermodynamics to statistical mechanics (SM) as his major extended and now canonical example. Nagel was keenly aware of the limitations of mechanics, mentioning its failure to explain electrodynamical phenomena, but also noting that other candidates for a universal theory have arisen. Additionally, Nagel addressed the epistemological (and metaphysical) problem often arising from successful reduction as to whether such reductions amounted to a repudiation of everyday folk categories, such as touch, tables, and headaches (on this, his answer was clearly "no"). From its inception, Nagel noted a

[4] The companion paper will present a more complete form of the reduction in section v, and similarly follows the Sommerfeld approach. However, the companion paper is additionally cross-referenced to several standard contemporary texts on optics, electromagnetic theory, and quantum electrodynamics. See my "The Explanation of Optics by Electromagnetic Theory: A Virtually Nagelian Reduction," available in draft form on my website at http://www.pitt.edu/~kfs/, and later to be submitted for publication.

distinction between what he termed "homogeneous" reductions, where no novel properties were introduced, and "heterogeneous" reductions, in which macroscopic properties like "temperature" were reduced to mechanical quantities associated with molecules. It was the latter types of reductions that drew most of Nagel's attention. Though Nagel stressed he was mainly interested in *theory* reduction, he often speaks of reduction of a s*cience* or a branch of science. This seems to be mainly an elliptical way of referring to the theories *and* the aggregated *experimental laws*, however, though there is some discussion of the role of auxiliary theories and borrowed laws in his 1961 chapter.

It is useful to situate the general approach of Nagelian reduction within the general category of "explanations." It is not possible to find textual evidence that Nagel was specifically generalizing the Hempel-Oppenheim Deductive-Nomological (DN) model (some call this the "Popper-Hempel" model) since the original, 1949 publication of the Nagel model contains no bibliographic references.[5] But the 1961 version places reduction within the context of explanation, and explanation itself has four patterns according to Nagel, the first and oldest of which is the deductive model.[6] Nagel emphasized the Aristotelian antecedents of this pattern in his chapter two. And Nagel did write in 1961 that "reduction, in the sense in which the word is here employed, is the *explanation* of a theory or a set of experimental laws established in one area of inquiry, by a theory usually though not invariably formulated for some other domain" (my italics).[7]

The Nagelian account of theory reduction is commonly summarized by presenting his two conditions of connectability and derivability, as well as noting that there were additional nonformal requirements for a successful reduction (though these often go unmentioned in the literature). Less noted is the fact that the connectability and derivability conditions were only developed after several *other* formal conditions were presented. Nagel repeatedly asserted the first formal condition is that the "assertions, postulates, or hypotheses of each of the sciences are available in the form of *explicit statements*, whose meanings are assumed to be fixed in terms of the procedures and rules of usage appropriate to each discipline" (my italics).[8] This is an "elementary"

[5] Possibly what initially motivated the deductive aspect of the Nagel model was his perceived need to accomplish "the derivation of the Boyle-Charles Law for ideal gases from the assumptions of the kinetic theory of matter." Nagel, "Meaning of Reduction," p. 109.

[6] Nagel, *The Structure of Science*, p. 21.

[7] *Ibid.*, p. 338.

[8] Nagel, "Meaning of Reduction," p. 112, and compare Nagel, *The Structure of Science*, p. 345.

sine qua non (later on I refer to this as the "condition zero" of reductions), but in discussions of reduction, with some exceptions, it has largely been ignored as the rational reconstruction and axiomatization of theories has gone out of philosophical fashion. And though often pursued in physics, such explicit codification is very rare in biology.[9] The approach to implementing the "explicit statements" condition in Nagel was via classical logical empiricist theory structure, involving theoretical and observational distinctions, as well as "coordinating definitions (or rules of correspondence)" associating theoretical with observational terms, but not necessarily confined to the first-order predicate calculus. Nagel's second formal condition was the related point, already noted, that the terms in both reduced and reducing sciences have codified established usage and meanings. Nagel cited the old issue of the definability of theoretical terms in observational language, but preferred to permit the possibility of the autonomous "theoretical" character of terms like "temperature." It is only within his "third formal consideration" that Nagel introduces what he now characterizes as "the formal requirements that must be satisfied for the reduction of one science [*sic*] to another." These are the well-known conditions of derivability and connectability. The condition of derivability stipulates that "the experimental laws of the secondary [reduced] science (and if it has an adequate theory, its theory as well) are shown to be the logical consequences of the theoretical assumptions…of the primary [reducing] science."[10] The connectability assumption is a corollary, in a sense, of this derivability condition, since, as Nagel noted, if the secondary science has a term "A" not in or simply definable in the primary science, additional "assumptions" must be introduced to logically permit the reduction. Restating this as "two necessary formal conditions," Nagel wrote that:

> (1) Assumptions of some kind must be introduced which postulate suitable relations between whatever is signified by 'A' and traits represented by theoretical terms already present in the primary science. The nature of such assumptions remains to be examined; but without prejudging the outcome of further discussion, it will be convenient to refer to this condition as the "condition of connectability." (2) With the help of

[9] A classic example is Hans Reichenbach, *Axiomatization of the Theory of Relativity* (Berkeley: California UP, 1969). In 1937 Joseph H. Woodger axiomatized and formalized genetics, but this had little influence on the philosophy of biology; see his *The Axiomatic Method in Biology* (Cambridge, UK: University Press, 1937). Interestingly, I found that to analyze scientific change (not reduction), I had to axiomatize two theories in immunology—see chapter 5 of my *Discovery and Explanation in Biology and Medicine* (Chicago: University Press, 1993).

[10] Nagel, *The Structure of Science*, p. 352.

> these additional assumptions, all the laws of the secondary science, including those containing the term 'A,' must be logically derivable from the theoretical premises and their associated coordinating definitions in the primary discipline. Let us call this the "condition of derivability."[11]

In his 1961 chapter, Nagel then turned to a discussion of the nature of the connectability assumptions, rejecting their status as logical connections but remaining fairly neutral as between thinking of them as conventions or as factual or material statements requiring empirical evidence for their warrant. Though in his 1979 essay he also briefly reformulated the convention interpretation within an instrumentalist approach, the consistently preferred theme found in his writings is to treat the connectability assumptions as factual claims. In the 1979 essay, he also used the term "bridge laws," which by then had become more widely employed in the philosophical literature. And in this later paper, he also suggested that we distinguish between how properties or predicates are analyzed, utilizing an extensional approach for them, and sketched how entities were to be treated. For entities, he opined that interpreting them as identity statements was reasonable, thus following earlier suggestions by Sklar, Causey, and me (references below).[12] Also in his 1979 article, Nagel reconsidered the nature of these connections in the light of a strongly critical analysis by Feyerabend, but curiously he did not refer to the similar and more generally influential ideas raised by Kuhn in his *Structure of Scientific Revolutions*.[13] Nagel strongly challenged Feyerabend on the "incommensurability" issue, arguing that two theories could not be inconsistent if they were incommensurable. Nagel also disagreed with Feyerabend's views that all scientific vocabulary, including observational terms, were globally infected by the theory in which they functioned. Nagel's incisive criticisms are well worth reviewing in detail for those who still subscribe to these Feyerabendian—and Kuhnian—positions. In 1979, Nagel was also more explicit about the implications of the use of approximations and simplifications in logical derivations, but I will say more about this refinement in later sections.

[11] *Ibid.*, pp. 353–54.

[12] Interestingly, in his 1949 essay, "The Meaning of Reduction in the Natural Sciences," p. 111, Nagel refers to an "identification" of temperature and mean kinetic energy, but places the term in quotes, and in his subsequent analysis opts for nonidentification—see text above.

[13] Paul Feyerabend, "Explanation, Reduction, and Empiricism," in Herbert Feigl and Grover Maxwell, eds., *Scientific Explanation, Space, and Time*, Minnesota Studies in the Philosophy of Science III (Minneapolis: Minnesota UP, 1962), pp. 28–97; and Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago: University Press, 1962).

II. THE RESPONSE TO NAGEL'S REDUCTION MODEL: 1960–2000

As just noted, Nagel's reduction model drew early and extensive criticism from Feyerabend, and less directly but equally importantly from Kuhn in 1962. The topic thus quickly became ripe for further analysis, one account pursued by Sklar in his 1964 Princeton Ph.D. dissertation and his 1967 article,[14] and another by myself in my Columbia dissertation and my 1967 essay.[15] My account drew extensively on Nagel but also on other analyses of reduction, including work by Woodger, Quine, and Suppes, but explicitly attempted to incorporate the critiques of Popper, Feyerabend, and Kuhn into a more pragmatic model of reduction.[16] This general reduction model focused on the derivability of a secondary or reduced theory which only had close analogies with the original secondary theory but was a corrected version in the reduced domain—corrected with the assistance of the reducing theory and its associated experiments. Because close analogy is not formally and precisely definable, some writers on reduction have seen this modification as problematic; more about this later. In my dissertation, I also applied Nagel's analysis to optics and electromagnetic theory, to valence theory and quantum chemistry, and to genetics in biology. In later papers, principally in a 1977 article, I also further generalized the model by incorporating a means for treating the replacement of an earlier theory with a reduction of the experimental domain of the previous science, as well as permitting "partial reductions"—more on this point just below. The 1977 generalization was both in accord with Nagel's original model permitting reduction of the "experimental laws" of an earlier theory which had been repudiated (or was inadequate), and was able to accommodate historical examples such as phlogiston and aether theories. This tack also had been pursued by Kemeny and Oppenheim in their alternative to Nagel's account.[17]

[14] Lawrence Sklar, "Intertheoretic Reduction in the Natural Sciences," unpublished Ph.D. dissertation, Princeton University, 1964; and Sklar, "Types of Inter-Theoretic Reduction," *British Journal for the Philosophy of Science*, XVIII, 2 (August 1967): 109–24.

[15] In the fall of 1961, Nagel suggested to me that an analysis of reduction, taking account of Feyerabend's recent critique of Nagel, as well as a then-forthcoming book by Kuhn on scientific revolutions, might be an interesting dissertation topic. Excellent advice!

[16] Kenneth F. Schaffner, "Approaches to Reduction," *Philosophy of Science*, XXXIV, 2 (June 1967): 137–47; and Schaffner, "The Logic and Methodology of Reduction in the Physical and Biological Sciences," unpublished Ph.D. dissertation, Columbia University, 1967.

[17] John G. Kemeny and Paul Oppenheim, "On Reduction," *Philosophical Studies*, VII, 1/2 (January–February 1956): 6–19.

I termed this 1977 generalization the "General Reduction Replacement" (GRR) model. The GRR model reintroduced the Nagel-like connectability assumptions, generalized to include corrections as well as direct connectability of a (revised) reducing theory $T_B(*)$ with T_R* as its condition one. The expression "$T_B(*)$" designated the possibility of a revised reducing theory if needed to accomplish the reduction of T_R* and T_R. The GRR condition two reiterated the Nagel derivability condition using the generalized connectability assumptions. Given the satisfaction of these conditions, a third requirement noted that T_R* corrects T_R in that T_R* makes more accurate predictions, or, if T_R was by then considered inadequate or "false" and thus "replaced," that $T_B(*)$ would make more accurate predictions in the "domain" of T_R. (The domain concept will be elaborated on more in section VI below.) Finally, the fourth condition of the GRR model stated that T_R is explained by $T_B(*)$ in the sense that T_R and T_R* are strongly analogous, and $T_B(*)$ indicates why T_R worked as well as it did historically, or that T_R's domain is explained by $T_B(*)$ even when T_R is replaced.

I noted in a footnote to condition one that "the distinction between entities and predicates will in general be clear in any given theory/model, though from a strictly extensional point of view the distinction collapses," an issue we shall return to later. I also indicated that such a GRR model has as a limiting case what I previously termed the "general reduction paradigm" (or "model") of 1967, which in turn yielded Nagel's model as a limiting case. Furthermore, the use of the weak sense of "or" in conditions one, two, and four allowed the "continuum" ranging from reduction as subsumption to reduction as explanation of the experimental domain of the replaced theory, with a "*partial reduction*" of some components of T_R also envisaged. I wrote that "to allow for such a continuum, T_R must be construed not only as a completely integral theory but also as a theory dissociable into individual assumptions, and also associated with an experimental subject area(s) or *domain(s)*." Examples of these replacements and partial reductions in both medicine and physics were sketched in my 1977 essay.[18] A main example there was that "*portions* of optical *theory* and the theory's explanations of experiments are preserved and modified in the context of a relativistic construal of Maxwell's electromagnetic theory."[19] These are important topics to which I return in section VI.

[18] Schaffner, "Reduction, Reductionism, Values, and Progress in the Biomedical Sciences," in Robert G. Colodny, ed., *Logic, Laws, and Life: Some Philosophical Complications* (Pittsburgh: University Press, 1977), pp. 143–72, at pp. 148–51.

[19] *Ibid.*, p. 151.

My sketch of a modified Nagel model applied to genetics drew a series of critiques by David Hull, who argued that Mendelian-Morganian genetics was related to molecular genetics in a complex many-many relation, and that the earlier theory was in effect being replaced and not reduced by molecular genetics. I responded to these critiques,[20] with those replies to Hull reprised in my book in 1993, pp. 437–45. Other philosophers of biology joined this debate; perhaps the most influential contribution was Kitcher's 1984 critique of Mendelian reduction, followed by similar arguments from Rosenberg.[21] Sarkar contributed a nuanced book on reductionism in biology, including a critique of quantitative behavioral genetics, but was critical of theory reduction.[22] Critiques and responses to this antireductionist analysis in genetics were formulated by Stent and Waters, as well as myself.[23] In 1973, Thomas Nickles distinguished two approaches to reduction, suggesting that in addition to the Nagel-Schaffner type, recovering a previous theory by taking certain quantities to limits such as zero or infinity was another sense of reduction.[24] Batterman further developed this approach, and Butterfield more recently addressed it in new ways.[25] More formal as well as set theoretic accounts of reduction were also developed by Stegmuller and Sneed and by Balzer and Dawe.[26]

Another line of criticism of the Nagel approach developed out of Wimsatt's critique suggesting we consider not theories but mechanisms, and in parallel out of Salmon's shift from a statistical relevance

[20] Schaffner, "Reductionism in Biology: Prospects and Problems," in R. S. Cohen, ed., *PSA 1974: Proceedings of the 1974 Biennial Meeting of the Philosophy of Science Association* (Dordrecht, Netherlands: Reidel, 1976), pp. 613–32.

[21] Philip Kitcher, "1953 and All That: A Tale of Two Sciences," *Philosophical Review*, XCIII, 3 (July 1984): 335–73; and Alexander Rosenberg, *The Structure of Biological Science* (New York: Cambridge, 1985).

[22] Sahotra Sarkar, *Genetics and Reductionism* (New York: Cambridge, 1998).

[23] Gunther S. Stent, "Promiscuous Realism," *Biology and Philosophy*, IX, 4 (October 1994): 497–506; C. Kenneth Waters, "Genes Made Molecular," *Philosophy of Science*, LXI, 2 (June 1994): 163–85; and Schaffner, *Discovery and Explanation in Biology and Medicine*, chapter 9.

[24] Thomas Nickles, "Two Concepts of Intertheoretic Reduction," this JOURNAL, LXX, 7 (Apr. 12, 1973): 181–201.

[25] Robert W. Batterman, *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence* (New York: Oxford, 2002); and Jeremy Butterfield, "Emergence, Reduction and Supervenience: A Varied Landscape," *Foundations of Physics*, XLI, 6 (June 2011): 920–59.

[26] Joseph D. Sneed, *The Logical Structure of Mathematical Physics* (Dordrecht, Netherlands: Reidel, 1971); Wolfgang Stegmüller, *The Structure and Dynamics of Theories* (New York: Springer-Verlag, 1976); W. Balzer and C. M. Dawe, "Structure and Comparison of Genetic Theories: (I) Classical Genetics," *British Journal for the Philosophy of Science*, XXXVII, 1 (March 1986): 55–69.

model of explanation to a causal mechanical process approach to explanation.[27] Craver, drawing in part on Salmon, began a response, further developed in the influential Machamer-Darden-Craver article (MDC), which has led to the burgeoning literature on "mechanisms," along with side critiques on the Nagel and Nagel-like models.[28] Dealing with the mechanism literature and its relations to reduction would take us beyond the scope of this article, but see my introduction to these topics.[29]

One variant, or possible competitor, of the Nagel-Schaffner approach has been termed the "New Wave" of reduction. This approach followed on the Churchland and Hooker analyses, and was subsequently defended by Patricia Churchland and John Bickle.[30] The core modification made in the "New Wave," following Churchland and Hooker, was the claim that in place of dealing with a corrected reduced theory T_R*, an "image" or restricted part of T_B is constructed "completely within the framework and vocabulary of T_B" which is equivalent to T_R*.[31] The "New Wave" was the subject of a searching critique by Endicott in which he summarized the key similarities and differences between my account and the Churchland-Hooker model.[32] Endicott concluded that the "New Wave" approach is tantamount to adopting a replacement and *not* a reduction perspective (but only my 1967 article is cited and not the more developed 1977 essay). I would agree with Endicott's dismissal of the "New Wave" approach, and add that though replacement would be consistent with a broadly eliminativist theme, which has been pursued by both Churchlands in many publications, it makes no clear sense as an approach to the broader notion of *reduction*. Furthermore, if a substantive close analogical relation is

[27] Wesley C. Salmon, *Scientific Explanation and the Causal Structure of the World* (Princeton: University Press, 1984); and William Wimsatt, "Reductive Explanation: A Functional Account," in Cohen, ed., *PSA 1974*, pp. 671–710.

[28] Carl F. Craver, "Beyond Reduction: Mechanisms, Multifield Integration and the Unity of Neuroscience," *Studies in History and Philosophy of Biological and Biomedical Sciences*, XXXVI, 2 (June 2005): 373–95; and Craver, *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience* (New York: Oxford, 2007).

[29] Schaffner, "Reduction: The Cheshire Cat Problem and a Return to Roots," *Synthese*, CLI, 3 (August 2006): 377–402.

[30] John Bickle, *Psychoneural Reduction: The New Wave* (Cambridge: MIT, 1998); Bickle, *Philosophy and Neuroscience: A Ruthlessly Reductive Account* (Boston: Kluwer, 2003); Paul M. Churchland, *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind* (Cambridge: MIT, 1984); Patricia Smith Churchland, *Neurophilosophy: Toward a Unified Science of the Mind-Brain* (Cambridge: MIT, 1986); Clifford Hooker, "Towards a General Theory of Reduction [Parts I–III]," *Dialogue*, XX, 1–3 (March, June, and September 1981): 38–59, 201–36, 496–529.

[31] Bickle, *Philosophy and Neuroscience*, p. 17; also see Bickle, *Psychoneural Reduction.*

[32] Ronald P. Endicott, "Collapse of the New Wave," this JOURNAL, XCV, 2 (February 1998): 53–72.

accepted as part of the "New Wave" analysis, then that account is indistinguishable from my model. Endicott does not accept my model, however, because he believes that it is still open to other antireductionist objections, especially to the more general "multiple realizability" problem, to which I now turn.

One recurring criticism of Nagelian kinds of reduction is the multiple realizability (MR) argument, which itself has had a number of forms. MR has also been specifically directed against Nagelian connectability by Hull in connection with the Schaffner variant and by Kitcher (1984) in the more Nagelian form. But in its most general form, the MR objection is not only directed against Nagelian-type reductions but against reducibility in general, with special emphasis on the reducibility of "functional" mental entities, properties, and processes to physical ones. This at least is the path that the Fodor-Putnam-Kim approach takes. The general intuition of MR seems to be that cross-level relations are either impossible to state or, if stateable, are too complex and open ended. They also destroy unity of explanations through extensive heterogeneity, and do not have the character of appropriate "laws" since they include "or" connectives in them. These arguments are said to hold not just for the more complicated biological sciences, and especially for the psychological sciences, but also are said to apply to physical high-level properties such as temperature, and also presumably to light (though I have not seen any arguments on this point in the philosophical literature).

In 1999, Kim advanced an alternative to the Nagel analysis which Kim termed a "functional model of reduction."[33] This model was intended to overcome what Kim termed the "philosophical emptiness of Nagel reduction" and to incorporate aspects of emergence. That model explicitly incorporates MR to an extent, but without additional development the functional model appears to be seriously flawed. According to Batterman's argument, it cannot rationally accommodate the extensive explanatory diversity its account of MR permits.[34] Further below I look at alternative ways that MR can be incorporated within Nagel-like models.

Generally, bridge laws need not conform to a preference for simple laws or expressions, such as are found in physics and chemistry; they need only work satisfactorily as part of reductions, a point I return to in the following section. In reductions in physics, the connectability assumptions may well turn out to be relatively simple.

[33] Jaegwon Kim, "Making Sense of Emergence," *Philosophical Studies*, xcv, 1/2 (August 1999): 3–36.

[34] Batterman, *op. cit.*, chapter 5.

They are not so in biology, but this does not argue against successful reduction, though some "boggled skeptics" may think it does. In his incisive but often-overlooked 1999 critique of MR, Sober pointed out that there is no clear reason to fear disjunctive laws, offering several examples where they are quite acceptable in terms of confirmability and utility.[35] That such disjunctions can be "open ended" is for Sober no argument about in-principle reducibility, and in point of fact, actual cases will depend on the sciences involved. The MR thesis has, in addition, been readdressed even more systematically in recent years, as noted in the following section. As Sober asks: "Are we really prepared to say that the truth and lawfulness of the higher-level generalization is *inexplicable*, just because the…derivation is peppered with the word 'or'?"[36]

III. THE RESPONSE TO NAGEL'S REDUCTION MODEL IN THE TWENTY-FIRST CENTURY

Given the extended and mostly negative criticism of Nagelian-type reductions over the past four decades, only some of which is summarized in the previous section, it is somewhat remarkable that the past several years have seen a number of articles generally defending Nagelian types of reductions.[37] In particular, two very recent and lengthy articles by prominent philosophers of science have strongly defended this kind of reduction, essentially in its classical form.[38] Both articles do make some use of arguments criticizing the critics of Nagelian reduction that appeared in the 1990s, but assemble those arguments and refute them with additional forceful defenses of classical theory reduction. The first, an essay from Hartmann's group appearing in *Erkenntnis* in 2010, was provocatively entitled

[35] Elliott Sober, "The Multiple Realizability Argument against Reductionism," *Philosophy of Science*, LXVI, 4 (December 1999): 542–64.

[36] *Ibid.*, pp. 552–53.

[37] Peter Fazekas, "Reconsidering the Role of Bridge Laws in Inter-Theoretical Reductions," *Erkenntnis*, LXXI, 3 (November 2009): 303–22; Colin Klein, "Reduction without Reductionism: A Defence of Nagel on Connectability," *The Philosophical Quarterly*, LIX, 234 (January 2009): 39–53; Paul Needham, "Nagel's Analysis of Reduction: Comments in Defense as Well as Critique," *Studies in History and Philosophy of Modern Physics*, XLI, 2 (May 2010): 163–70; Rasmus G. Winther, "Schaffner's Model of Theory Reduction: Critique and Reconstruction," *Philosophy of Science*, LXXVI, 2 (April 2009): 119–42.

[38] Foad Dizadji-Bahmani, Roman Frigg, and Stephan Hartmann, "Who's Afraid of Nagelian Reduction?" *Erkenntnis*, LXXIII, 3 (November 2010): 393–412; and Butterfield, *op. cit.* Both of these articles were published with companion articles that relate to more technical and applied examples of reduction: Dizadji-Bahmani, Frigg, and Hartmann, "Confirmation and Reduction: A Bayesian Account," *Synthese*, CLXXIX, 2 (March 2011): 321–38; and Butterfield, "Less Is Different: Emergence and Reduction Reconciled," *Foundations of Physics*, XLI, 6 (June 2011): 1065–135.

"Who's Afraid of Nagelian Reduction?" and systematically defended what they termed the "Generalized Nagel-Schaffner Model of Reduction" or GNS. The second, Butterfield's essay in 2011 on "Emergence, Reduction, and Supervenience: A Varied Landscape," had a larger domain of analysis in its scope, but the core of that paper also notes that the author "favour[s] the Nagel-Schaffner account" against both the "New Wave" critics and virtually all of the published critiques of this classical approach, which are analyzed in detail in this paper.

In this section, it will be most useful to group some of the major points that have developed in the literature of the past five years under the two salient headings of derivability and connectability, using these recent articles.

The central thesis in the Nagel model of reduction was the derivability of the reduced theory from the reducing theory, albeit with the needed connectability assumptions. Derivability was noted by Nagel to be "both necessary and sufficient for reduction," inasmuch as "the condition of derivability obviously entails connectivity."[39] In his 1979 essay, Nagel suggested that "good approximations" to explicit derivations were satisfactory, noting Galileo's law of falling bodies and Kepler's laws as reduced by Newtonian mechanics. Moreover, Nagel added, that is "actual scientific practice," where there are "simplifications and approximations of various kinds" usually involved in deriving a law from a theory. Here Nagel cited the law of the simple pendulum, adding that this kind of derivation is a proper kind of deduction.

Nagel did not in his 1979 essay refer to my 1967 article, in which a rigorous deduction could be made of a modified reduced law or theoretical assumption from the reducing theory. In that article, I acknowledged that what was deduced only bore a "close similarity" with the original reduced theory and that "the relations between old and corrected reduced theories should be one of strong analogy—that is…they possess a large 'positive analogy'."[40] As already indicated, this pragmatic proposal of a "close analogy" has occasioned strong criticism of this variant of the classical Nagel model, as well as recent vigorous defense of the appropriateness of the "close analogy" terminology. Some of this was mentioned above in the discussion of the "New Wave" approach. However, a more recent objection to this proposal is Winther's article on my model. There, Winther correctly identifies the "informal" "analogous" element, as in condition four of the GRR account, but he misinterprets it as needing a formal rendition. In practice, scientists work fairly easily with strongly

[39] Nagel, *The Structure of Science*, p. 355n5.
[40] Schaffner, "Approaches to Reduction," p. 144.

analogous conceptions, for example, in relating aether-based optics and relativity, as shown in section v below. Second, Winther's critique of the complexity of relating corrected Mendelian genes and molecular genes had largely been previously answered by me.[41]

Thus, the issue of the *prima facie* vagueness and nongeneral character of the "strongly analogous" requirement in the various Schaffner models has drawn frequent criticism. However this is not the case, oddly enough, for recent strongly formally oriented philosophers of science. Hartmann's group, which is among the more formally directed contributors, is quite comfortable with context-specified strong analogy (in their analysis with the relation between corrected thermodynamics derivable from statistical mechanics and classical thermodynamics), and with broadly case-based determination of that requirement, noting that "we should not expect there to be a general theory of analogy."[42] Amplifying on this theme, Dizadji-Bahmani et al. summarize then criticize the varied objections to the "close analogy" notion, such as being too vague and arbitrary, as lacking a general characterization of the nature of analogy, and as assuming that the reduced theory as originally conceived has in reality been replaced rather than reduced. This set of concerns, I noted earlier, putatively favors the "New Wave" approach, but, as I did, Dizadji-Bahmani et al. find them without merit. Dizadji-Bahmani et al. add, further defending the notion of a close analogy, that T_R and T_R* must share "all essential terms," indicating that this is an important constraining requirement that illuminates the notion of close analogy. They also reformulate the prediction criterion of the GRR (condition three), but relax it to one that says that T_R* should "be *at least equally empirically* adequate as T_R." This seems reasonable, though perhaps only a rhetorical change.[43]

[41] See Schaffner, *Discovery and Explanation in Biology and Medicine*, pp. 437–87, on connectability assumptions, including responses to Hull and Kitcher, which are not adequately noted by Winther in his "Schaffner's Model of Theory Reduction: Critique and Reconstruction." Winther does propose an additional intriguing thesis in his paper, namely that "A reconstructed Schaffnerian model could…shed light on mathematical theory development in the biological sciences and on the epistemology of mathematical practices more generally." *Ibid.*, p. 119. But pursuing this notion would take me beyond the scope of this article.

[42] Dizadji-Bahmani, Frigg, and Hartmann, "Who's Afraid of Nagelian Reduction?" p. 409.

[43] In my *Discovery and Explanation in Biology and Medicine*, chapter 5, I embraced the notion of "empirical adequacy" as one of the primary criteria for theory assessment. I defined empirical adequacy as "the ability of a theory to explain empirical results, whether these be singular reports of experimental data or empirical generalizations such as Snell's law in optics," and noted furthermore that "precision generally is a species of empirical adequacy, and scope and consilience are also species of empirical adequacy."

Butterfield, another philosopher of formal orientation, has also recently commented on the issue of approximate derivability as well as on the "close analogy" problem. He maintains that in any given case of reasonable reduction, we can both deduce a proposition of interest (a fact or a law) that is "approximately true; and that we can quantify how good the approximation is."[44] Butterfield, like Dizadji-Bahmani et al. and some of the earlier commentators whom Butterfield cites on this point, including Nickles, writes "(a) we should not expect, and (b) we do not need, any such general account" of close approximation or strong analogy. In addition, he adds "in reduction, T_B need only imply an approximation or 'cousin' of T_R, corrected by the lights of T_B," a view which "fits well with a central claim of the companion paper" on reduction by Butterfield, with its extensive application of reduction, especially to limiting cases.[45]

Nickles as well as Wimsatt originally suggested that the contested term "strong analogy" might find some clarification in what Batterman later called the "physicist's sense of reduction."[46] This is the notion that we can recover a reduced theory by taking some quantity in the reducing theory to a limit, often 0, but sometimes ∞. Batterman developed this idea in considerable detail in his 2002 book as well as in more recent writings. This notion of what occurs at certain extreme limits has also been explored in depth by Butterfield and Norton.[47] As we will see in the following application section, these various strategies can be seen as broadly consistent with a Nagel-like approach, and as part of the approximations and simplifications invoked as part of the derivability claims. That said, these strategies may fit better into a *partial* reduction approach rather than a strongly systematic and fully rigorous general theory approach to reduction. More to come on this below.

Finally, regarding the derivability condition, this approach, especially as relaxed yet still constrained as in the paragraphs above, should not be conflated with the problems with the Hempel model of DN explanation. Some writers appear to subscribe to this red herring.[48] The long debate about DN explanation, amounting to a half century of writings by both proponents and critics, remains fertile

[44] Butterfield, "Emergence, Reduction and Supervenience," p. 939.

[45] *Ibid.*, p. 939, and Butterfield, "Less Is Different," generally.

[46] Batterman, *op. cit.*

[47] Butterfield, "Less Is Different"; and John D. Norton, "Approximation and Idealization: Why the Difference Matters," *Philosophy of Science*, LXXIX, 2 (April 2012): 207–32.

[48] Craver, "Beyond Reduction"; but see my *Discovery and Explanation in Biology and Medicine*, pp. 286–88, for a defense of deduction as the only recognized form of inference that is truth preserving.

ground for discussions about explanation,[49] but the theory reduction model discussed in this section carries its own strengths and limitations, best addressable in the context of reduction examples.[50]

Let us now turn from deducibility issues to recent discussions of connectability assumptions. There have been several lines of criticism, some related to Nagel-like theory reduction, and some addressing more general concerns of reducibility, including the much-discussed multiple realizability (MR) arguments noted in the previous section. Recall from our first section that Nagel had struggled with the nature of his connectability assumptions from his 1949 essay through to his 1979 article. In general he tended to interpret these assumptions as empirical hypotheses, though that might be conditional on the nature of the explicit formulation chosen. Also in his 1979 paper, Nagel favored an extensional interpretation of predicates. In 1968, I also independently developed an extensional analysis for predicate connections.[51] By 1979, Nagel seems to have accepted the by then almost consensus definition, defended earlier by Sklar, Causey, and myself,[52] that important entity statements were synthetic identities, such as a light wave being identical to an electromagnetic wave. Butterfield in his 2011 paper strongly defends co-extensions for both entities and predicates as quite powerful and sufficient for Nagelian types of reductions.

In contrast, however, Dizadji-Bahmani et al. argue that there is an important difference between entity identity statements, which they view as *internal* or part of the reducing theory, and bridge laws for properties, which they contend are *external* to the reducing theory. Using the canonical SM example, they write that "there is nothing in the kinetic theory of gases per se that tells us to associate mean kinetic energy with temperature."[53] But this internal/external distinction seems fallacious for two reasons. First, the entity identities

[49] See the extensive discussion of this theme in Kitcher and Salmon, eds., *Scientific Explanation*, Minnesota Studies in the Philosophy of Science XIII (Minneapolis: Minnesota UP, 1989).

[50] Jim Woodward has suggested (personal communication) that the derivability condition may be necessary but not sufficient. This seems right, but the Nagelian reduction models discussed in this paper do have additional constraints, including informal conditions that may serve as sufficiency conditions for reductive explanations.

[51] Schaffner, "The Watson-Crick Model and Reductionism," *British Journal for the Philosophy of Science*, xx, 4 (December 1969): 325–48.

[52] Sklar, "Intertheoretic Reduction in the Natural Sciences"; Schaffner, "Approaches to Reduction"; and Robert Causey, *The Unity of Science* (Dordrecht, Netherlands: Reidel, 1977).

[53] Dizadji-Bahmani, Frigg, and Hartmann, "Who's Afraid of Nagelian Reduction?" p. 404.

are not themselves part of or internal to the reducing theory, but are added to the theory, per se, just as are the property association laws. Second, both types of bridge laws, or more accurately connectability assumptions, require empirical support of the assumptions, and a strong case can be made that extensional equivalence for predicate terms is the correct approach. Dizadji-Bahmani et al. differ on this point and prefer to view the predicate relations more generally. Arguments for the former position, however, have been made by Butterfield, and an empirical example involving optics and electromagnetic theory will be outlined in section v of this article.

The past year has also seen further analysis of the MR issue, quite pointedly as a defense of Nagel-Schaffner forms of reduction. On this discussion, see Dizadji-Bahmani et al. as well as Butterfield's account of MR. Both essentially agreed with Sober's 1999 critique discussed earlier.

IV. PARTIAL OR "DAPPLED" REDUCTIONS—THE BACKGROUND

The past 40-plus years of explication of reduction in science and applications of reduction ranging from physics through biology and to neuroscience have yielded a few general truths, one of which is that actual reduction is hard to do. Serious attempts at the more interesting inhomogeneous reductions in the special sciences, including genetics and the cognitive neurosciences, have uncovered immense complexity including redefinitions of fundamental concepts such as the gene,[54] and the articulation in consciousness studies of what still appears to be a very "hard problem."[55] In physics, the situation is considerably simpler, with much of the analysis and application in the reduction literature still driven by Nagel's original SM exemplar. But even in the case of this example, deeper analysis by Feyerabend, Sklar, Batterman, Callendar, Dizadji-Bahmani et al., and Butterfield (involving phase transitions in SM), among others, has shown that there are contentious issues with both connectability and derivability relations.[56] Some writers, such as Kim, have claimed that in point of

[54] Peter Beurton, Raphael Falk, and Hans-Jörg Rheinberger, eds., *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*, Cambridge Studies in Philosophy and Biology (New York: Cambridge, 2000).

[55] David John Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (New York: Oxford, 1996).

[56] Craig Callender, "Reducing Thermodynamics to Statistical Mechanics: The Case of Entropy," this JOURNAL, XCVI, 7 (July 1999): 348–73; Sklar, *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics* (New York: Cambridge, 1995); also see Feyerabend, *op. cit.*; Batterman, *op. cit.*; Butterfield, "Emergence, Reduction, and Supervenience"; and Dizadji-Bahmani, Frigg, and Hartmann, "Who's Afraid of Nagelian Reduction?"

fact there are no valid Nagelian reductions,[57] whereas others still hope for a systematic general reduction even in the most complex of sciences, human consciousness.[58]

As a compromise between these extremes of "no reductions exist" and fairly simple Nagel-type reductions, in 2006 I proposed that though there are *very close analogues* of such systematic and sweeping reductions in the physical sciences—one is developed in the next section—they are quite rare and depend on special requirements, and in the end also turn out to be partial reductions. In the biological sciences and the neurosciences, however, these attempts at sweeping reductions, when pressed, tend to fade away like the body of the Cheshire Cat, leaving only a smile. Those Cheshirean "smiles" that remain are essentially fragmentary or patchy explanations, also describable as "creeping" reductions or sometimes as "local" reductions. Elsewhere I have provided an example involving a simple organism's feeding behaviors.[59] I will return to the analysis of partial reduction at the close of this paper.

This idea that there still may be very close analogues of Nagelian *systemic* reduction, however, suggests that we might consider in more detail another physics example that has been mentioned by Nagel, Sklar, and myself as meeting the (modified) Nagel conditions for theory reduction. It is likely that the reduction literature has overly focused on the canonical SM example, and an additional virtually systematic reduction may well cast new light on this type of reduction. This additional example is the reduction of physical optics by electromagnetic theory, subsequently referred to in this article as the "optics" example. That example, which has apparently never been extensively presented in the published literature on reduction, will actually support many of the original Nagelian claims, if in some cases slightly modified. But it will also show that even in this case the reduction is ultimately, if at the margins, a *family of often-approximate reductions*, most which work quite well as a systematic theory reduction, but others which fail completely for some parts of the target domain. The take-home lesson from this example, also applicable to the canonical SM example, is that even in physics, reductions are ultimately partial or, in Butterfield's term, "local," though familially

[57] Kim, "Emergence: Core Ideas and Issues," *Synthese*, CLI, 3 (August 2006): 547–59.

[58] Patricia S. Churchland, "Commentary," in Stephen Marcus, ed., *Neuroethics: Mapping the Field* (New York: Dana Press, 2002), p. 56.

[59] Schaffner, "Reduction: The Cheshire Cat Problem and a Return to Roots"; and Schaffner, "Etiological Models in Psychiatry: Reductive and Nonreductive," in Kenneth S. Kendler and Josef Parnas, eds., *Philosophical Issues in Psychiatry: Explanation, Phenomenology, and Nosology* (Baltimore: Johns Hopkins, 2008), pp. 48–90.

related within a domain. Thus reductions have a "dappled" nature, to use a felicitous term introduced by Cartwright, albeit from a strongly pluralist and nonreductionist stance.[60]

V. THE OPTICS-ELECTROMAGNETIC THEORY (OPTICS) EXAMPLE

A detailed and systematic account of exactly how the optics reduction example works can be conveniently and relatively succinctly found in two back-to-back books by the distinguished physicist Arnold Sommerfeld.[61] Sommerfeld published six advanced textbooks in the 1940s covering all of physics, which were based on his extensive lectures on the topics delivered in the 1930s. Volume III was entitled *Electrodynamics*, and volume IV, *Optics*.[62] The optics in volume IV is developed reductionistically from Maxwell's theory as delineated in his volume III, and the two texts represent an in-depth extended exemplar of an almost-sweeping classical Nagelian reduction. This explication is also written largely in the Euclidean-Newtonian mode of entire subfields such as reflection and refraction, the optics of moving bodies, dispersion, and diffraction being mathematically derived from a small number of integrated universal physical laws supplemented with relatively simple connections between the fundamental terms in the reduced and reducing theories.

But such a comprehensive, sweeping, deductively developable account seems to be dependent on some rather stringent requirements, basically akin to Nagel's "first" formal condition of explicit formulation. Both reduced and reducing fields need to be representable in terms of a relatively small number of principles or laws, though Nagel was not committed to axiomatization in first-order predicate calculus. In addition, for classical Nagelian-type reductions, connections between the two fields need to be straightforward and relatively simple, though the connections may well be far from obvious. Both of these stringent conditions, simple axiomatizablity and simple connectability, fail in significant ways in more complex sciences such as molecular genetics and neuroscience, though that they do fail, or would fail, was not necessarily

[60] Nancy Cartwright, *The Dappled World: A Study of the Boundaries of Science* (New York: Cambridge, 1999).

[61] Other authoritative optics texts covering much of the same ground are Max Born, Emil Wolf, and A. B. Bhatia, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th (expanded) edition (New York: Cambridge, 1999); and Geoffrey Brooker, *Modern Classical Optics* (New York: Oxford, 2003).

[62] Arnold Sommerfeld, *Lectures on Theoretical Physics*, vol. III, *Electrodynamics* (New York: Academic, 1950); Sommerfeld, *Lectures on Theoretical Physics*, vol. IV, *Optics* (New York: Academic, 1950).

obvious at the beginning of the Watson-Crick era of comparatively uncomplicated molecular biology.

That Nagel's two conditions do fail is, however, an *empirical* question and not decidable by philosophers in advance of the empirical research. In the example I pursue in the following pages, the notion of "simple" as a modifier of axiomatizability and connectability will need to be altered in subtle ways, and following from those modifications the correlate *of simple deducibility* will also require some adaptation, albeit of the kind that Nagel himself accepted in his 1979 article. The example as delineated below is of necessity presented as four compressed "fragments" of the more complete reduction of optics by electromagnetic theory.[63] As Nagel himself noted when he first introduced the SM example in 1949, it "is not possible, without producing a treatise on the subject, to exhibit the complete argument," adding that he would "therefore fix my attention on a small fragment of the complicated analysis, the derivation of the Boyle-Charles Law for ideal gases from the assumptions of the kinetic theory of matter."[64] Similarly, in his more expanded 1961 treatment, Nagel noted that he would only be able to deal with "but a small part" of this example, again just the derivation of the ideal gas law.[65] In the optics case, we happily have that "treatise" at hand, in the two Sommerfeld volumes.

Turning back to more specific issues related to this example, in his *Optics* volume Sommerfeld begins his preface noting that "[t]his volume is closely connected with 'Electrodynamics,' Vol. III of my lectures." He adds that "Not only the formalism of Maxwell's equations but also their intrinsic character, the invariance with respect to the group of Lorentz transformations, is adopted from Vol. III and is assumed to be known."[66]

Sommerfeld's first sentence expresses the recurrent theme of his treatment of optics, which is to provide explanations of broad classes of optical phenomena, such as reflection and refraction, the optics of moving media and sources, dispersion, crystal optics, and diffraction, in terms of electrodynamics. Sommerfeld does so either by beginning with Maxwell's equations and developing the explanations deductively, or by interpreting the light vector **V** in optical wave theory in terms of the **E**, or electric force vector, but then proceeding deductively from an older and simpler form of wave optics (a shortcut

[63] See footnote 4 for a reference to a more complete account of this reduction example.
[64] Nagel, "Meaning of Reduction," p. 109.
[65] Nagel, *The Structure of Science*, p. 343.
[66] Sommerfeld, *Optics*, p. v.

of sorts, but with the Maxwell equations at the ready if necessary). (As is conventional, the bold-faced type indicates a vector quantity.) But Sommerfeld's second sentence also implicitly reveals that the original Maxwell equations of the 1860s and '70s, even in their classical simplified form as rendered by Hertz in 1890, will be modified by newer contexts and by additions to the classical theory. Lorentz's electron theory of the 1890s was an augmentation and improvement of Maxwell's theory, and the Lorentz transformations published in 1904 are *not* interpreted by Sommerfeld as Lorentz originally did, but in a symmetrical Einsteinian manner.[67] Furthermore, for some quantum phenomena, principally the photoelectric effect but also some forms of radiation phenomena, the classical Maxwellian theory is only part of the "truth" and needs to be "complemented" by quantum particle optics, as I will indicate in more detail below.

Sommerfeld's general reductive strategy as just summarized is evident in his treatment of the classical subdomains of optics. Here I briefly summarize the introductory parts of Sommerfeld's analysis for four representative subdomains of physical optics: reflection and refraction, diffraction, the optics of moving bodies, and the quantum theory of light including the photoelectric effect.[68] Generally, Sommerfeld proceeds from the four basic equations of Maxwell in their later (1890) simplified Hertz-like form. These equations are: $\mathrm{div}\,\mathbf{D} = \rho$, $\mathrm{div}\,\mathbf{B} = 0$, $\varepsilon_0\partial\mathbf{E}/\partial t = \mathrm{curl}\,\mathbf{H}$, and $\mu_0\partial\mathbf{H}/\partial t = -\,\mathrm{curl}\,\mathbf{E}$, as well as $\mathbf{D} = \varepsilon_0\mathbf{E}$ and $\mathbf{B} = \mu_0\mathbf{H}$. Here, **E** is the electric field strength, **B** the magnetic field induction, **D** the (di)electrical displacement vector, and **H** the magnetic field strength. The constants ε_0 and μ_0 represent electrical and magnetic permeabilities of the vacuum. (Readers might note that Feynman suggests that the **D** and **H** vectors are just hidden ways of referring to what is going on in media, and that **E** and **B** are the fundamental quantities in Maxwell's theory.[69])

[67] See Schaffner, "The Lorentz Electron Theory of Relativity," *American Journal of Physics*, XXXVII, 5 (May 1969): 498–513; and Robert Rynasiewicz, "Lorentz's Local Time and the Theorem of Corresponding States," in Arthur Fine and Jarrett Leplin, eds., *PSA 1988: Proceedings of the 1988 Biennial Meeting of the Philosophy of Science Association* (East Lansing: Philosophy of Science Association, 1988), pp. 67–74.

[68] Of the coverage in classical optics texts, this mainly leaves out only the topics of crystal optics, since polarization is treated under reflection and refraction, and interference is dealt with via diffraction. Contemporary texts add such topics as lasers and quantum computing.

[69] Richard P. Feynman, Robert B. Leighton, and Matthew Sands, *The Feynman Lectures on Physics*, vol. II (New York: Basic, 2010), pp. 32–34.

In his volume on *Electrodynamics*, Sommerfeld derived the wave equation for electromagnetic propagation from mathematical manipulations of equations. That wave equation, in its simplified plane wave form proceeding along the x axis in empty space and **E** vibrating in the y direction, can be expressed as $\partial^2\mathbf{E}_y/\partial t^2 = \varepsilon_0\,\mu_0\partial^2\mathbf{E}_y/\partial x^2$. Here the constant product $\varepsilon_0\,\mu_0$ has the dimensions of an inverse velocity squared, which can be calculated from purely electromagnetic measurements and equals $\sim 3 \times 10^8$ meters per second—the velocity of light. This is the more modern version of the famous 1862 derivation which led Maxwell to conclude that light is an electromagnetic wave, though in Maxwell's original derivation this result was obtained on the basis of a very complex electromagnetic aether model. Thus "light" is plausibly identical with "an electromagnetic wave," though this potential identification is as yet *incomplete*. This identity or connectability assumption is only the preface to the more refined identity that then guides the remainder of Sommerfeld's analysis. That refined identity is an identification of the traditional "light vector," called **V**, with the electric force vector **E**, and *not* with the magnetic force vector **H** (or equivalently the magnetic induction **B**).

In Hertz's experiments, recounted so brilliantly in his introduction to his collected papers *Electric Waves*, Hertz supported the identification of light with electromagnetic waves (though not yet that **V** = **E**). This latter accomplishment was left to Wiener's work in 1890, described below, and in point of fact Hertz added an 1891 note in his *Electric Waves* referring to Wiener's "beautiful experiments." However, the important Hertz experiments using his spark coil resonator and detector demonstrated that the *properties* of electromagnetic waves are strongly analogous and likely fully equivalent to those of light waves. These properties included the electromagnetic waves' reflection, interference, diffraction, and refraction.

To determine whether it is **E** or **H** that is the light vector **V**, Sommerfeld writes that it is necessary to examine what occurs when light affects a photographic plate and produces an image. In this situation, Sommerfeld notes that "an electron is removed from a silver bromide or chloride molecule and thereby a silver molecule is prepared to blacken during the development of the film. Only the electric field strength **E** can accomplish this." Note that this analysis does not explicitly appeal to a photon-electron action or to known physiology, though Sommerfeld conjectures that "Since, moreover, the processes occurring in the eye's retina are quite similar [to photography] (both phenomena are without doubt 'photoelectric effects') we have good reason to give the name 'light vector' to the field vector **E** rather than to the magnetic vector **H**."[70] This argument is more strongly and empirically supported by the elegant 1890 experiments of Weiner on the photographic process.

The gist of Weiner's experiment involved writing the equations for a light wave that is reflected from a silver mirror and impinges on a photographic plate tilted at an angle. The analysis of Wiener's experiment (following Sommerfeld's account) then begins from the wave equation for the electric vector components, and obtains expressions for the specific locations of the nodes and antinodes, contingent on the additional initial conditions of the experimental apparatus. The observations obtained by Wiener were complete confirmations of the predicted blackening by **E** nodes, and Sommerfeld writes: "Thus the electric vector **E** is indeed photographically active and is to be considered the light vector. The magnetic vector is *not* the light vector. Its antinodes alternate with those of the electric vector, the first being on the surface itself."[71] These "beautiful experiments," to use Sommerfeld's term echoing Hertz's admiration for Weiner's experiments, indicate quite clearly how connectability assumptions are established in reductions, and that they are synthetic empirically based identity statements.

With the connectability established (here an identification of **E** with **V**), Sommerfeld can then explain many of the received laws of optics by derivations from Maxwell's fundamental equations or

[70] Quotes are from Sommerfeld, *Optics*, p. 56. On this point also see Brooker, who writes regarding a choice of **E** or **H** as the light vector **V** that "the reason for this choice does not lie in electromagnetism which treats **E** and **H** (or **B**) more or less symmetrically, but in atomic physics: atoms are usually much more interested in the **E** field of light than in the **B** field, because their strongest transitions are electric-dipole transitions." Brooker, *op. cit.*, p. 6.

[71] Sommerfeld quotations are from his *Optics*, p. 58; also see his p. 58 for a diagram of Wiener's apparatus. For the original experiment, see Otto Wiener, "Stehende Lichtwellen und die Schwingungsrichtung polarisirten Lichtes," *Annalen der Physik und Chemie*, CCLXXVI, 6 (1890): 203–43. An excellent analysis of Wiener's paper, with an explanation of the nature of standing waves and Wiener's experimental apparatus, is also online at the blog: http://skullsinthestars.com/2008/05/04/classic-science-paper-otto-wieners-experiment-1890/ (accessed March 23, 2012) run by the pseudonym Dr. SkySkull, an associate professor of physics. Wiener's experiment was also discussed by Pierre Duhem as the first motivating experiment supporting the famous D-thesis, later usually called the "Quine-Duhem thesis." But Duhem's discussion of Wiener was seriously limited, commenting only on the aspects of the Wiener experiment that affected the Neumann polarization theory, and not on Maxwell's theory; see pp. 184–86 of Pierre Duhem, *The Aim and Structure of Physical Theory* (Princeton: University Press, 1954), translated by P. P. Wiener from the second French edition published in 1914. Duhem was notoriously critical of Maxwell, which may account for Maxwell being ignored. A fuller treatment of this issue regarding Duhem, Wiener, and Maxwell can be found in my companion paper mentioned in footnote 4 above.

from equations that represent **E** waves.[72] **H** is not disregarded, however, since specifying needed boundary conditions for many of the derivations also requires knowledge or postulation of the values of both **E** and **H** at interfaces (in reflection and refraction) or apertures (in diffraction).

The first laws Sommerfeld obtains are the laws of reflection and refraction (Snell's law). These are derived using premises for the **E** wave equation incident at angle α, proceeding at angle β in the refracted media, and the boundary conditions at the interface of a plane surface with two different dielectric (ε) and magnetic constants (μ) (for example, air_1 and water_2). Sommerfeld writes Snell's law based on his derivation as:

$$\sin\alpha/\sin\beta = \sqrt{\varepsilon_2\mu_2} \,/\, \sqrt{\varepsilon_1\mu_1} = n_{12}$$

where n is the relative index of refraction (here between air and water).

This derivation is exact, but that prediction for actual substances is incorrect because of complications such as infrared resonance vibrations and the fact that Maxwell's theory in its original, simpler form does not account for dispersion (though it can be extended to do so as in Sommerfeld's chapter III).[73]

Sommerfeld next turns to derivations of Fresnel's famous sine and tangent laws for the ratio of the relative amplitudes of incident, reflected, and refracted polarized light. In this case he obtains a somewhat more complicated (more general) form of the ratios which upon simplification yield Fresnel's famous sine expression for the intensities of reflected and refracted light waves where the electric vector is perpendicular to the plane of incidence. He also derived another expression for Fresnel's tangent law.[74] Thus both of the Fresnel intensity laws are actually simplifications and involve approximations, in comparison with the more complex and rigorous results derivable using Maxwell's equations.[75]

[72] I should add here that this order of presentation is not exactly Sommerfeld's, who has to wait until he has clarified the nature of **V** (he eventually does this on pp. 56–57 of his *Optics*), but he proceeds to explain reflection and refraction beginning on p. 6, albeit focusing on the amplitude of the **E** fields.

[73] Reflection and refraction laws are derived using the *wave* equations, not the simplified ray diagrams sometimes employed. Brooker cautions us to always think in terms of waves, with rays as normals to the wave fronts, but also adds that ray diagrams are much better (far less cluttered) for graphical representations. Brooker, *op. cit.*, p. 12.

[74] These Fresnel ratios have played an important role in structural realism discussions in philosophy of science—see John Worrall, "Structural Realism: The Best of Both Worlds?" *Dialectica*, XLIII, 1–2 (June 1989): 99–124.

[75] See Julius Stratton, *Electromagnetic Theory* (New York: McGraw-Hill, 1941), pp. 492–94. Also compare John D. Jackson, *Classical Electrodynamics*, 3rd ed. (New York: Wiley, 1999), pp. 302–06.

Similar, though considerably more mathematically complex, approximations are introduced in my second "fragment" of Sommerfeld's optics, this one on *diffraction* or the bending of light waves around an obstacle including a straight edge. The general point which Sommerfeld makes is that, as in refraction results, most of the calculations of diffraction results involve approximations and simplifications, in contrast to a rigorous and complete derivation.

The topic of diffraction in optics and more rigorous treatments of diffraction have been the subject of analysis in advanced electromagnetic theory texts, and have also received recent attention in the philosophy of science literature by Saatsi and Vickers.[76] Suffice it to say, however, that the Sommerfeld approach described briefly above, which is essentially equivalent to the Kirchhoff 1883 optical approach, is explained and can be *further corrected* by the more rigorous electromagnetic analysis *based on Maxwell's equations*, including extending the analysis to vector (not only scalar) representations, as well as using Fourier analysis to represent the multiple frequency light waves encountered in practice. However these rigorous solutions are *quite limited*, requiring special conditions and special mathematics (for details especially see Sommerfeld's *Optics*, section 38)—restrictions that indicate the limited direct application of Maxwell's fundamental equation, somewhat along the lines of Cartwright's views on this topic.[77]

Here are three salient examples from diffraction theory, the first two dealing with connectability problems and the third with a derivational issue. First, there are certain properties, such as the traditional black or opaque screen found in optics, which cannot even be defined in Maxwellian theory, though one can work around this issue for most experimental cases.[78]

A second connectability issue arises in attempts to solve a diffraction problem exactly using Maxwell's equations. In general, the diffracting object needs to satisfy "Maxwell's equations both outside and inside the…object" as well as "the proper boundary conditions on the surface of that object."[79] But these requirements are *not physically realizable* for light in the case of the famous circular disk and "Poisson's spot," analyzed earlier in optics by Fresnel, Poisson, and Arago. Finally, third, Sommerfeld expresses frustration about

[76] Juha Saatsi and Peter Vickers, "Miraculous Success? Inconsistency and Untruth in Kirchhoff's Diffraction Theory," *British Journal for the Philosophy of Science*, LXII, 1 (March 2011): 29–46, at p. 35.

[77] Cartwright, *How the Laws of Physics Lie* (New York: Oxford, 1983).

[78] Sommerfeld, *Optics*, p. 205.

[79] *Ibid.*, p. 247.

applying Maxwell's theory to the classical problem of the rainbow's colors and structure, writing that the sophisticated spherical harmonic series and specialized Bessel functions needed here "converge so slowly that they become practically useless."[80] This example indicates the limitations of the derivational application of Maxwell's theory to a phenomenon clearly in the *domain* of optics but not solved adequately by any theory of optics up to that time. Thus what we have in virtually all these diffraction examples are at best approximations and perhaps more correctly close analogues of experimental results that are not rigorously derivable from the reducing theory.

For other examples of attempts at more modern rigorous accounts see Brooker and Jackson, and also Saatsi and Vicker's discussion, the latter with applications to the issue of scientific realism.[81] These analyses indicate that something quite like the Schaffner reduction *correction*, also involving close analogy, is present in this example. In point of fact, Brooker's analysis using Maxwell's equations clarifies why the widely accepted diffraction analysis of Kirchhoff "worked as well as it did historically" (to recite condition four of the Schaffner GRR model above). Brooker's detailed reasons can be found in his text, and he includes a reference to a Fourier analysis of the differences between the Kirchhoff-derived (call it "T_R") and Maxwell-derived (here "T_R*") treatments of diffraction. Brooker concludes that the comparison indicates "there are good reasons why we can get away with using Kirchhoff boundary conditions at a diffracting aperture."[82]

To introduce our third example from optics, chapter II of Sommerfeld's optics text analyzes the optics of moving media, including the velocity of light, aberration, Fresnel's "partial drag," and the famous Michelson-Morley experiment's attempt to detect the "aether wind." These latter two examples, Fresnel's partial-drag expression and Lorentz on the Michelson-Morley experiment, show (here *kinematically* and *not* dynamically) how a reducing theory can explain both experimental results obtained under earlier theories and why those theories worked as well as they did, but by *discarding some key aspects of those theories (the aether)* and correcting and reinterpreting those key results.

Our fourth and final optics example relies on the concluding section of Sommerfeld's optics of moving media chapter, this titled "The Quantum Theory of Light." There Sommerfeld begins from

[80] *Ibid.*, p. 248.

[81] Brooker, *op. cit.*; Jackson, *op. cit.* Especially see Brooker, pp. 70–72, and Saatsi and Vickers, *op. cit.*

[82] Brooker, *op. cit.*, p. 72.

what he terms Einstein's "radical" 1905 paper on the photoelectric effect, experimentally discovered by Hertz in 1897. Sommerfeld notes this involves "a new elementary particle, the photon, that is a return in a sense to the older Newtonian corpuscular theory of light." This is, he adds, "an extremely remarkable situation" "from an epistemological point of view."[83] Though some phenomena can be treated from either the wave or particle perspective, "the photon theory, at least in its present state of development is unable to account precisely for polarization and interference phenomena. *Therefore we are forced to adopt a dualistic conception of light*: not Huygens *or* Newton, *but* Huygens *and* Newton."[84] Sommerfeld, however, then refers to Bohr's notion of *complementarity*, which he sees as a better characterization of light than *duality*. In introducing the photoelectric effect, Sommerfeld explicitly recognizes the limitations of the traditional Maxwellian-Lorentzian components, and even the Einsteinian relativistic components, of traditional optics. This underscores the patchiness or incompleteness of this reduction, albeit perhaps more at its margins, and even though the Maxwellian theory continues to serve as the basis of the explanation of broad domains of optics.

Above, I have briefly summarized how Sommerfeld proceeds by deduction, approximation, and simplified shortcuts using equations from wave optics. Though the Sommerfeld account of optics, largely based on classical and relativistic electromagnetic theory, holds in the main, and by allowing approximations of various sorts covers broad domains of optics exceedingly well, it clearly fails at the margins. This is the case especially in connection with the photoelectric effect, and often at places closer to the core, as in the case of the black-screen and circular-disk examples. The account above, thus, carries with it a deeper lesson—all reductions, even relatively "simple" paradigm reductions in physics, are most likely partial and thus not fully systematic.[85] That point of view and how to analyze it in more general terms will be taken up in our last section below.

[83] Sommerfeld, *Optics*, p. 87.

[84] Sommerfeld still believed this was the case in 1949, based on the Preface date of his *Optics* book. Why Sommerfeld did not acknowledge the power of quantum electrodynamics in the Dirac and later forms is an interesting historical question that will be addressed in my companion paper; see note 4, also note 90 below.

[85] It is of interest that Butterfield also seems to agree with this view when he writes, "there is surely no single best sense of 'reduction'." And if nondisjunctiveness is required, that will "unquestionably make for reductions narrower in scope. What really matters, scientifically and philosophically, is to assess, in any given scientific field, just which reductions hold good, and how narrow they in fact turn out to be." Butterfield, "Emergence, Reduction and Supervenience," p. 942.

VI. THE CONDITIONS FOR A PARTIAL REDUCTION

To further clarify this notion of a partial reduction, which is the *prima facie* case even in the putatively systematic optics example outlined above, let us consider how the conditions for such a type of reduction might be formulated. The general conclusion (and guiding intuition) of this section is that the GRR model can function not only as a touchstone of systematic reduction, but can also be an appropriate framework for specifying where such systematic reductions *fail*, and thus become partial reductions.

Now a partial or incomplete reduction might be "partial" in several somewhat different senses, with a "core" sense indicating the reduction is incomplete in some important respect. Even when its reductions are "partial," however, they may still reduce extensive topic areas. Recall that Nagel sometimes equivocated whether his analysis of reduction was focused on *theories* only, or whether in some case the elements within the reduction relation might be *sciences* or branches of a science. In the brief remarks that follow, I address both foci: theories and science, conceiving of the latter as a combination of the theories, the observational generalizations, and phenomena in that "domain" (more on this below).[86]

Raising the question whether an inter*theoretic* reduction is partial or incomplete presumes that a theory can be distinguished into component parts. This assumes some explicit codification, much as Nagel did with his "first" formal condition. Recall Nagel's requirement that the "assertions, postulates, or hypotheses of each of the sciences are available in the form of *explicit statements*, whose meanings are assumed to be fixed in terms of the procedures and rules of usage appropriate to each discipline." Making such postulates "explicit" will often need to be achieved by the scientific peers in the subject domain, as Hertz did for Maxwell's theory, though philosophers with a taste for rational reconstruction and axiomatization could fulfill this role.

Physics, like mathematics, has historically been able to articulate theories based on a small number of related hypotheses that have extraordinary scope of application. We saw this in the example of Maxwellian electrodynamics in the previous section, but also encounter it in classical mechanics, thermodynamics, and quantum mechanics. The bottom line, then, is that there will be a somewhat different

[86] Here and further below the term "phenomena" is used in the Bogen and Woodward sense; see James Bogen and James Woodward, "Saving the Phenomena," *Philosophical Review*, XCVII, 3 (July 1988): 303–52. I leave open the interesting question whether special instruments, such as the Michelson interferometer, should also be included in the "domain," as Kuhn included them in his expansive notion of "paradigm."

flavor to the nature of the "partial" character of reductions in terms of scope and approach in different sciences. Much of the analysis in this section is oriented toward partial reductions in *physics*. Several other publications of mine and others treat reductions in biology, where causal mechanical temporal models are analyzed.[87]

Because of the neater and often-axiomatized theory structures found in physics, there we also often have the possibility of identifying a specific core component hypothesis of a putative reduced theory that has failed to be explained by the reducing theory. Three types of such reductive failures suggest themselves. First, we may consider those situations in which a core hypothesis (including what may be termed a "law") in the potentially reduced theory is *not* derivable, in spite of general success in deriving other core assumptions and results of the reduced theory. Examples were discussed in the optics example and included the photoelectric effect, the photon, and quantum results more generally. Relatedly, in the canonical SM example, failure to adequately derive the second law of thermodynamics has generated considerable discussion.[88] A second type of failure of a reduction would be the inability to derive key experimental results (phenomena) that are part of the putatively reduced theory's standard and successful applications. We encountered these in several diffraction examples in the optics case, such as the circular disk and the black screen, where we noted these were at best weaker analogous derivations. Third, there are experimental phenomena that fall under the "domain" of the reduced theory, which even the reduced theory does not adequately explain, such as the photoelectric effect and the rainbow, but which constrain the adequacy of the putative reducing theory.

In the optics area analyzed above, Sommerfeld employed as his main reducing theory earlier versions of Maxwell's electromagnetic theory as reformulated by Hertz, but as added to by Lorentz, who introduced electrons and a force law, as well as the famous Lorentz contraction and the Lorentz transformations. Additionally, Lorentz's interpretation of the "contraction" and his asymmetric characterization of the transformations were significantly recast in Einstein's special theory of relativity, which Sommerfeld also accepted as part of his main reducing theory. What we seem to have here, then, is a

[87] The background to different approaches to reductions in physics and biology is developed in chapter 9 of my *Discovery and Explanation in Biology and Medicine*, with more detail on reductions in biology presented in my "Reduction: The Cheshire Cat Problem and a Return to Roots."

[88] See references in footnote 56 above.

series of reducing theories, with each of the later ones absorbing major parts of the previous theory, *but not all* of the previous reducing theory. Starkly put, the aethers of Fresnel, Kirchhoff, Maxwell, and Lorentz are rejected and replaced by relativistically invariant Maxwellian fields in the Einsteinian theory. Thus, there arises the need to provisionally fix the form (and the time period) of the reducing theory before using it to explain (or determining a failure to explain) the reduced theory—or parts of the reduced theory. In Sommerfeld's case, the reducing theory was the mainstream electrodynamics of 1935–1945. This introduces an important *diachronic* feature into reductions, but it can be handled by specifying which form or codification of the reducing theory is being utilized, as just noted. That said, one of the *limitations* (that is, partial failure) of the reduction at any given time period may *subsequently* be removed, and an incomplete feature of the putative reduction repaired, if and when a new, more powerful successive theory is developed and applied to the reduced domains. This was the case for light quanta, achieved by Dirac's theory of quantum electrodynamics, but only as later augmented by Feynman, Dyson, and Schwinger in the late 1940s.[89,90]

As briefly noted in section III, the notion of the possibility of a "partial" reduction also arose earlier in the summary of the GRR approach in terms of a continuum of reduction relations in which T_B (or T_B*) can participate in reducing T_R and T_R*. To allow for such a continuum, T_R needed to be construed not only as a completely integral theory but also as a theory dissociable into weaker versions of the theory, and also appropriately associated with an experimental subject area(s) or domain(s). This consideration suggests that we might appeal in some manner to the GRR to further specify and systematize our notion of a partial or incomplete reduction.

[89] In *Discovery and Explanation in Biology and Medicine*, chapter 5, I outlined the structure of diachronic change of the immune response theories, which I termed temporally "extended theories." There I appealed to several levels of abstraction, in which there was constancy at high abstraction levels and changes of more specified assumptions at lower abstraction levels. I have not attempted to apply this tack to electrodynamics in the present paper, but I do so in my companion article; see footnote 4.

[90] It is as yet unclear to me why Sommerfeld never referred to Dirac's (or any other form of) quantum electrodynamics. Possibly this is because that theory was in considerable trouble by the mid-1930s, and did not resolve the difficulties of various key quantities becoming infinite until renormalization was introduced in the late 1940s and 1950s. See Julian Schwinger's preface for an account of these developments in his *Selected Papers on Quantum Electrodynamics* (New York: Dover, 1958). I hope to resolve this in my companion article; see footnote 4.

As a general model, then, I suggest that we conceive of partial reductions as largely completed reductions containing exceptions or failures of attempted reductions. I now frame this within the previously discussed GRR model, viewed as a touchstone for characterizing successful reductions but here extended to also cover partial theory replacements. The notion of "partial failure" used here indicates that such a failure is *comparatively minor* in comparison with the reductive success in *most areas* of the reduction. A metrical analysis of this "comparatively minor" aspect of failure is not feasible, at least at this point, though generally a review of standard texts and review articles in the subject areas of actual scientific reductions will indicate the nature, scope, and significance of any incompleteness in the reduction. (As pointers to such incompleteness and reductive failures, look for strong disagreements—Kuhnian "crises"—over "anomalies," including anxieties over specific concepts, laws, and key experiments.) In the light of the discussed need for codifiability, however, to be added to those four GRR conditions of connectability and derivability, correction, and strong analogy, is a requirement for such a codification. This we might term a "condition zero" that reformulates what Nagel in his 1961 account called his "first formal" condition, requiring that those structures involved in a reduction relation be made explicit. This might read:

> (0) The set of theory(ies)$\{T\}$ and/or the (branches of) science, comprised of $\{T\}$ and $\{O\}$ (experimental results and generalizations) and $\{D\}$ (domain phenomena) is sufficiently codified (but not necessarily axiomatized) that there are effective procedures for reaching a determination whether the following four GRR conditions for a successful reduction hold, or whether their partial failure in the case of partial reductions is found.

From this perspective, partial reductions occur in the context of almost-complete reductions if and only if one or both of the generalized conditions of connectability (condition one) and derivability (condition two) of the GRR model partially fail, or if the corrections and close-analogy conditions (conditions three and four) partially fail. An example of the condition of generalized connectability failing occurred with the photon concept and the experimental photoelectric effect for Maxwell's theory, even when supplemented with Lorentz's electrons and interpreted relativistically. Thus, there the reduction is partial, in spite of the fact that most of optics is reduced by the modified Maxwell theory. Other partly experimental examples were types of diffraction failures, as in the nonconnectability (definability) of a black screen and the physical unrealizability of an infinitely thin but opaque circular disk in Maxwell's theory applied to optics. Should the condition of generalized *derivability* fail (recalling that the sense of "derivation" here does allow for approximations and close analogies between T_R* and the older T_R), then the reduction is also partial. This may occur, in spite of putatively adequate connectability assumptions holding, when the mathematical *derivations* are (at present) unsatisfactory, such as in cases where mathematical series of spherical harmonic and Bessel functions will not converge in diffraction examples. Relatedly, the recalcitrant rainbow example is also a derivational failure of Maxwellian reduction, but here in the experimental "domain" of optics and not in a reduced theory.

I believe that this set of considerations captures what we tend to find in most instances of reduction in the sciences, and that it represents what we encountered in the optics example as well. The optics case was and still is a most important sweeping kind of reduction, similar to the sweep of the SM example, though it is a partial reduction at the margins, absent an extension of the reducing theory to encompass quantum electrodynamics (and still remain the "same" reducing theory).

VII. SUMMARY AND CONCLUSION

In this article I have reviewed Ernest Nagel's three important essays that delineated the Nagel model of theory reduction, then examined many of the amplifications and revisions of kindred models of reduction. The account takes place over a time span of over sixty years, since Nagel's first article appeared in 1949. I argued that though some classically sweeping Nagelian reductions putatively can be found in physics, a closer inspection of one in-depth case of optics and electromagnetic theory indicates that in spite of its stunning successes, this reduction fails at the margins and perhaps even more centrally (for quantum phenomena). Even where reasonably successful, that reduction depends on approximations and in some cases on "work arounds" in order to accomplish its explanations. Though space restrictions prohibit a similar consideration of the canonical Nagelian thermodynamics-statistical mechanics example, it is likely that similar failures can be identified in that example. The optics example also indicates that closer attention should be paid to the specific temporal form of an evolving set of reducing theories, and the question when a theory change is sufficient to say that it is a "different theory" merits further attention. Many of these reductive themes were situated in the framework of the Nagel-like general reduction replacement (GRR) model;

however, many of these themes also indicate important nonformal and pragmatic aspects of what occurs in real scientific examples.

As the review above indicates, the Nagel model is still going strong, in spite of a vast array of critical analysis directed at it, and can play an important role in an explication of partial reductions. But though strong and still extensively discussed and applied, the Nagel model now appears in a more conditionalized and nuanced pragmatic context, one that reflects the more complex and "dappled" analyses of science and scientific change than could have been anticipated in the middle of the twentieth century. But the Nagel model and its variants have weathered the philosophical and scientific criticism, are alive and well, and seem very likely to continue to generate philosophical arguments and counter-arguments for many years to come.

KENNETH F. SCHAFFNER

University of Pittsburgh

NEW BOOKS

DREYFUS, HUBERT and SEAN DORRANCE KELLY: *All Things Shining: Reading the Western Classics to Find Meaning in a Secular Age.* New York: Free Press, 2011. xi, 254 p. Cloth: $26.00.

DWORKIN, RONALD: *Justice for Hedgehogs.* Cambridge: Harvard, 2011. xi, 506 p. Cloth: $35.00.

EARDLEY, PETER S. and CARL N. STILL: *Aquinas: A Guide for the Perplexed.* New York: Continuum, 2011. vii, 156 p. Cloth $75.00, Paper $19.95.

EBBS, GARY: *Truth and Words.* New York: Oxford, 2011. xiv, 338 p. Paper: $35.00.

EHRING, DOUGLAS: *Tropes: Properties, Objects, and Mental Causation.* New York: Oxford, 2011. viii, 250 p. Cloth: $75.00.

ELBOURNE, PAUL: *Meaning: A Slim Guide to Semantics.* New York: Oxford, 2011. viii, 174 p. Cloth: $75.00, Paper: $22.95.

ELDER, CRAWFORD L.: *Familiar Objects and Their Shadows.* New York: Cambridge, 2011. xi, 210 p. Cloth: $85.00.

ELLIOTT, KEVIN C.: *Is a Little Pollution Good for You? Incorporating Societal Values in Environmental Research.* New York: Oxford, 2011. x, 246 p. Cloth: $65.00.

ENGLEBRETSEN, GEORGE and CHARLES SAYWARD: *Philosophical Logic: An Introduction to Advanced Topics.* New York: Continuum, 2011. ix, 198 p. Cloth: $90.00, Paper: $27.95.

ENOCH, DAVID: *Taking Morality Seriously: A Defense of Robust Realism.* New York: Oxford, 2011. xi, 295 p. Cloth: $75.00.

FAULKNER, PAUL: *Knowledge on Trust.* New York: Oxford, 2011. x, 216 p. Cloth: $55.00.

FAYE, EMMANUEL: *Heidegger: The Introduction of Nazism into Philosophy.* New Haven: Yale, 2011. xxviii, 436 p. Paper: $27.50.

FENVES, PETER: *The Messianic Reduction: Walter Benjamin and the Shape of Time.* Stanford: University Press, 2011. xiv, 312 p. Paper: $24.95.

FEYERABEND, PAUL: *The Tyranny of Science.* Malden, MA: Polity, 2011. xii, 153 p. Paper: $19.95.

FINGER, ANKE, RAINER GULDIN, and GUSTAVO BERNARDO: *Vilém Flusser: An Introduction.* Minneapolis: Minnesota UP, 2011. xxx, 176 p. Paper: $21.00.

FISCHER, JOHN MARTIN: *Our Stories: Essays on Life, Death, and Free Will.* New York: Oxford, 2011. vii, 184 p. Paper: $24.95.

FISHER, DAVID: *Morality and War: Can War be Just in the Twenty-first Century?* New York: Oxford, 2011. 303 p. Cloth: $45.00.

FLANAGAN, OWEN: *The Bodhisattva's Brain: Buddhism Naturalized.* Cambridge: MIT, 2011. xv, 264 p. Cloth: $27.95.

FLEISCHACKER, SAMUEL: *Divine Teaching and the Way of the World: A Defense of Revealed Religion.* New York: Oxford, 2011. x, 576 p. Cloth: $110.00.

FLORIDI, LUCIANO: *The Philosophy of Information.* New York: Oxford, 2011. xxx, 405 p. Cloth: $55.00.

FORSTER, GREG: *Starting with Locke.* New York: Continuum, 2011. xi, 159 p. Cloth: $75.00, Paper: $19.95.

FORSTER, MICHAEL N.: *German Philosophy of Language: From Schlegel to Hegel and Beyond.* New York: Oxford, 2011. xi, 350 p. Cloth: $85.00.

FORSTER, PAUL: *Peirce and the Threat of Nominalism.* New York: Cambridge, 2011. xii, 259 p. Cloth: $82.00.