LOGIC AND PHILOSOPHY

LOGIQUE ET PHILOSOPHIE

INSTITUT INTERNATIONAL DE PHILOSOPHIE

ENTRETIENS DE DÜSSELDORF

27 août — 1er septembre 1978

LOGIQUE ET PHILOSOPHIE

publié par

G.H. VON WRIGHT

1980

MARTINUS NIJHOFF PUBLISHERS

LA HAYE / BOSTON / LONDRES

INTERNATIONAL INSTITUTE OF PHILOSOPHY

SYMPOSIUM IN DÜSSELDORF

27 August — 1 September 1978

LOGIC AND PHILOSOPHY

edited by

G.H. VON WRIGHT

1980

MARTINUS NIJHOFF PUBLISHERS
THE HAGUE / BOSTON / LONDON

Distributors:

for the United States and Canada

Kluwer Boston, Inc.
160 Old Derby Street
Hingham, MA 02043
USA

for all other countries

Kluwer Academic Publishers Group
Distribution Center
P.O. Box 322
3300 AH Dordrecht
The Netherlands

Library of Congress Cataloging in Publication Data CIP

Main entry under title:

Logic and philosophy.

Half title: Logic and philosophy, logique et philosophie.
"International Institute of Philosophy, symposium in Düsseldorf, 27 August–1 September 1978."
English and French.
1. Logic—Congresses. 2. Philosophy—Congresses.
3. Modality (Logic)—Congresses. 4. Probabilities—Congresses.
I. Wright, Georg Henrik von, 1916– II. International Institute of Philosophy.
III. Title: Logique et philosophie.
BC51.L58 160 79-27403

ISBN-13: 978-94-009-8822-4 e-ISBN-13: 978-94-009-8820-0
DOI: 10.1007/978-94-009-8820-0

CONTENTS/TABLE DES MATIÈRES

PREFACE

The *Entretiens* of the Institut International de Philosophie for 1978 were held in connection with the World Congress of Philosophy in Düsseldorf, from August 27 to September 1. The theme of the *Entretiens* was Logic and Philosophy (*Logique et philosophie*). The undersigned, then President of I.I.P., was responsible for the planning of the programme.

The programme was designed to consist of four sections with the headings Classical and Intuitionist Logic, Modal Logic and its Applications, Inductive Logic and its Applications, and Logic and Epistemology. The aim was also to convey to philosophers who are not experts in logic an informative and representative impression of some of the main sectors of the vast and rapidly expanding field of philosophical logic. At the same time it was thought that this impression should not be conveyed in the form of a series of survey papers but through presentations and discussions of specific topics falling under the main headings mentioned above.

For each section a *rapporteur* was nominated to read a paper and an *interlocuteur* to comment on it. The programme chairman is grateful that he was able to engage a representative selection of front rank philosophical logicians to perform the various tasks. The papers and the comments are printed in this volume in the order in which they appeared in the Programme of the *Entretiens.*

The various sessions of the *Entretiens* were attended by a numerous audience of participants to the World Congress, in addition to participating members of the I.I.P. The discussions from the floor after each paper were lively and stimulating. No attempt was made to summarize the discussions for publication. With the two speakers' mutual agreement, however, a reply by Professor Stegmüller to the comments on his paper by Professor Suppes is printed here.

The Institut International de Philosophie is much indebted to the Organizing Committee of the World Congress and in particular to its President, Professor Alwin Diemer, for assistance and cooperation with the practical arrangements. I.I.P. is also grateful to the Academy of Finland for generous financial support in organizing the *Entretiens.*

Helsinki
November 1978

GEORG HENRIK VON WRIGHT

DAG PRAWITZ

(*University of Stockholm*)

INTUITIONISTIC LOGIC: A PHILOSOPHICAL CHALLENGE

There is a tendency in recent discussions to try to reconciliate the differences between intuitionistic and classical logic by saying that each logic has its interest and is correct from its own point of view when suitably interpreted. But this is to overlook the fact that there is a real conflict between two positions of which at most one can be correct: classical logic is set forth as a system of universally valid canons of reasoning and intuitionistic logic was formulated in response to Brouwer's criticism of the classical claim and as a revision of this allegedly erroneous classical logic. In our deductive practice, we take, in fact, a stand on this conflict; even if we try to reconcile the two logics, our actual reasoning will usually show a clear preference for one of them. In philosophy, furthermore, we should reflect on our deductive practice and should be able to formulate explicitly what kind of reasoning is correct.

The conflict between classical and intuitionistic logic raises not only the question which logic is correct but also the question how one argues about the correctness of a logical law. This general methodological question was especially made acute by the intuitionistic position, which is perhaps the first serious exception to the usual unanimity on fundamental logical laws.

It seems clear, as has been suggested by Michael Dummett, that a philosophical discussion of the issues raised by intuitionistic logic has to be carried out to a great extent in a theory of meaning. Among others, I shall discuss some ideas about the meaning of the logical constants suggested by Dummett and myself that seem relevant in this connection.

1. THE ISSUE

The most well-known controversy between classical and intuitionistic logic concerns the question whether certain uses of the law of excluded middle are warranted. The question is not whether there exist sentences A for which the assertion "A or not-A" is not true. Obviously, there exists no such sentence A since it would have to be both false and not false, *i.e.* both

G.H. von Wright (ed.), Logic and Philosophy, 1–10.

not-A and not-not-A would then have to hold, which is contradictory.

By an intuitionistically perfectly acceptable *reductio ad absurdum*, this reasoning shows that the assertion not-(A or not-A) is false, *i.e.* that for any sentence, it holds that not-not-(A or not-A). Hence, the intuitionistic rejection of the law of excluded middle brings with it a rejection of the law of double negation (or what comes to the same thing: *reductio ad absurdum* where one assumes the negation of what is to be proved); in fact it is easy to see by reasoning doubted by nobody that the two laws are equivalent.

Consequently, intuitionistic logic, at least as first formulated by Heyting, does not contain the negation of any assertion made in classical logic but simply does not assert certain principles such as the law of excluded middle or the law of double negation occurring in classical logic.

2. HOW TO DEBATE LOGICAL LAWS: PRACTICE VERSUS THEORY

The more general philosophical question, which the conflict between classical and intuitionistic logic gives rise to, *viz.*, the question how one supports or challenges an alleged logical law, is quite as interesting as the conflict itself.

A concrete inference is often justified by quoting a general rule of which the inference is an instance. And sometimes one argues for the correctness of a rule or logical law by deriving it from a more basic one. But this cannot be the only way of supporting a logical law since otherwise we should be unable to defend the most basic ones; the selection of them would be just a matter of choice not open to rational discussion.

Sometimes the opposite view is advocated; *e.g.*, Brouwer meant that a logical rule gets its validity from the fact that its instances are valid inferences. On this view, a logic gets its validity from the fact that it describes a correct deductive practice. However, it is then difficult to see how one can argue for the correctness of an individual inference. Clearly, one cannot accept a view that one deductive practice is as good as another.

Similar questions about the relation between human practice and the rules or theory for this practice are asked not only in logic but everywhere in human affairs, *e.g.*, in the moral sphere where it is discussed whether an action is good because it agrees with general moral principles or whether conversely these principles are correct to the extent that they describe good actions. The right view seems here to be the one advocated by, *e.g.*, Goodman and Rawls that neither practice nor theory has priority but that both support each other. Our theoretical reflections that result in general principles try to make our practice intelligible and may lead to revisions of the practice, but conversely the practice may be stronger than a proposed theory and the theory may be rejected as unable to account for actual practice. Our actual practice and our theories about it have to be modified until the two levels agree with each other in what Rawls calls a reflective equilibrium.

In the case of our deductive practice, the theoretical level contains not only logical laws but also principles about ontology, truth and meaning. The dispute between classical and intuitionistic logic has to be resolved by finding an equilibrium between our inferential practice and a comprehensive theory that includes all these principles and makes our practice intelligible.

3. THE CLASSICAL THEORY OF MEANING AND TRUTH

One of the philosophically most interesting aspects of the dispute between classical and intuitionistic logic is just the fact that it makes acute several important questions about truth and meaning. It is often held that the classical laws are valid in virtue of the meaning of the logical constants involved in the laws. This is of course quite reasonable in itself: if the laws are valid, they should be valid in virtue of this meaning. The question is what is here to be understood by meaning.

According to a widespread view, supported by philosophers like Frege, the early Wittgenstein, and Carnap, the meaning of a sentence consists in the condition for its truth; to know the meaning of a sentence is to know how the world has to be when the sentence is true, in short, to know the truth condition of the sentence. Since Tarski's definition of truth, it has furthermore been common to identify the truth conditions with the condition of adequacy that Tarski proposed for a definition of truth. For instance, we have the following truth conditions for disjunctions and negations: a sentence "A or B" is true if and only if A is true or B is true; a sentence not-A is true if and only if A is not true.

I shall call the position just indicated the classical theory of meaning and truth. It may be thought that this theory should support classical logic. By combining the truth conditions for disjunctions and negations, we get that a sentence "A or not-A" is true if and only if the truth condition of A obtains or does not obtain. Since this truth condition just expresses the meaning of the sentence, its logical validity follows if we furthermore assume the principle that a truth condition either obtains or does not obtain, independently of our means of knowing which case is the actual one. But this principle, which we may call with Dummett the platonistic principle of truth, must of course not be taken for granted in a discussion of the validity of the law of excluded middle.

It is only by combining what I have called the classical theory of meaning with the platonistic principle of truth that we are able to support the laws of classical logic. The classical theory of meaning is itself quite neutral with respect to classical and intuitionistic logic. One must say that the theory is indeed quite empty. Not that there is necessarily anything wrong with the principle that the meaning of a sentence is determined by its truth condition or with the Tarskian truth conditions, but it is an illusion to think that these principles are able to clarify the concepts of meaning and truth.

This is especially clear if we recall that the general form of the Tarskian truth condition of a sentence A is "A is true if and only if P," where P is the proposition expressed by A. As has been pointed out by Dummett, these equivalences cannot simultaneously determine the notion of truth and the meaning of the sentences in question. If all we know about the concept of truth is that its extension is a set S determined by a number of equivalences of the form "$A \in S$ if and only if P," then the information A is true, *i.e.*, belongs to this set S, if and only if P cannot possibly give information also about the meaning of A. For all we know, S may be, *e.g.*, the set of false sentences.

Furthermore, it is to be noted that even if we already know what truth is, the equivalence "A is true if and only if P" cannot inform us about the meaning of A, unless we know that the equivalence is a truth condition in Tarski's sense, *i.e.*, that P is the proposition expressed by A. In other words, the very notion of truth condition presupposes that we know what it is for a sentence to express a proposition, *i.e.*, it presupposes the notion of meaning.

The classical theory of meaning is thus too poor to cast any light upon the notions of meaning and truth, let alone to give any guide in resolving the conflict between classical and intuitionistic logic.

4. REQUIREMENTS ON A THEORY OF MEANING

An adequate theory of meaning obviously has to place the notion of meaning in a richer context. The main function of the concept of meaning is to explain communication. In general, we explain people's way of using and reacting to utterances by, among other things, assuming that they know or do not know the meaning of the utterances. In an adequate theory of meaning, we should thus be able to deduce a certain use of the language from the meaning of its expressions. In particular, from an assumption that a person knows the meaning of an expression should follow that he uses the expression in certain ways.

This requirement on a meaning theory is a way of understanding Wittgenstein's slogan "meaning is use." Dummett has suggested that a meaning theory should respect this slogan in a stronger way. He observes that a person may know the meaning of an expression explicitly or implicitly. The knowledge is explicit when the person can explain or state what the meaning is, which requires that he knows the meaning of some other expressions, namely those used in the explanation. Knowledge of meaning is therefore in the end implicit, and implicit knowledge of the meaning of an expression must manifest itself in the use of the expression, Dummett argues. To this one may object that although knowledge of the meaning of an expression must imply a certain use of the expression, the knowledge may not manifest itself completely in any finite totality of individual behaviour. However, even my weaker requirements seem fatal for classical logic as we shall see.

5. DUMMETT'S TWO ASPECTS OF THE USE OF A SENTENCE

What use of a language should be deducible from the meaning of its expressions? Dummett has suggested that there are two main features of the use of an expression: the rules governing in what situations the expression is appropriately uttered and those governing the appropriate responses or expectations that follow upon uttering the expression; schematically, the conditions for uttering the expression and the consequences of uttering it. If the expression is an assertive sentence, the first feature is the conditions for asserting it, *i.e.*, the grounds on which it is correct to assert it, and the second feature is determined by the conclusions that can be inferred from the sentence, its deductive consequences.

A person cannot be said to know the meaning of an expression without knowing both aspects of its use. But if meaning is identified with truth conditions it is difficult to see how knowledge of meaning is to imply knowledge of use; in fact, it is even unclear what is to be meant by knowing a truth condition, since explicit knowledge cannot be demanded.

The situation is quite different if we try to make the condition for correctly asserting a sentence a constitutive element of its meaning. To contrast the notion of truth of a sentence with the condition for its correct assertion, we may note that a sentence may be true, but nevertheless it may be incorrect to assert it in a given situation, *viz.*, if the person in the situation in question does not have sufficient grounds for asserting it. Hence, the condition for asserting a sentence is something which takes into account the situation in which the sentence is asserted to a much greater extent than the truth condition does. Furthermore, and this is the main point, there is no mystery about what is meant by knowing the conditions for correctly asserting a sentence since such knowledge certainly shows itself in the behaviour.

There is clearly a parallel difference between the notion of a consequence of a sentence and the condition for correctly inferring a conclusion from a sentence.

As pointed out by Dummett, the two aspects of the use of a sentence, the conditions for correctly asserting it and for correctly inferring a conclusion from it, must stand in harmony with each other. Roughly speaking, the conclusions that can be rightly inferred from a sentence must be only such ones as are guaranteed to hold when the conditions for asserting the sentence are satisfied. A lack of harmony would mean that by making an assertion, a person could create false expectations or could commit himself to something that he was unable to fulfil although he had observed the condition for the assertion. When discovered, such a lack of harmony has of course to be mended by changing some of the rules.

In order to construct a meaning theory that meets the requirement discussed in the preceding section, we may consider the possibility of identifying meaning with the two aspects of use described above. Or, since the two aspects have to stand in harmony with each other, it may be enough to identify meaning with one of them and derive the other. However, there are several questions about the exact form of these conditions and about their sufficiency for an adequate meaning theory.

6. ON THE FORMULATION OF A "PRAGMATIC" MEANING THEORY

I shall only briefly discuss some problems that we have to deal with when we try to formulate the condition for correctly asserting a sentence and the consequences of asserting it, *i.e.*, problems that an explicit meaning theory along the lines of the preceding sections have to deal with.

To state the conditions for correct assertions in mathematics is to say how sentences of different forms are proved, *i.e.*, what is counted as a proof of them. One may expect that Gentzen's introduction rules or Heyting's explanation of the notion of proof could then be used. We would then say that to prove, *e.g.*, a conjunction "A and B" we have to prove A and we have to prove B, or that a proof of "A and B" is a pair consisting of a proof of A and a proof of B. However, this surmise is not quite correct. Obviously, we sometimes prove a sentence "A and B" without first proving the two sentences A and B; we may first prove a more complex sentence from which we infer "A and B," and this is a perfectly legitimate ground for asserting "A and B"; *nota bene* provided that the proof in question is a real proof.

The conditions for correct assertions thus seem to require that we already have a general notion of proof, but on the other hand, whether something is a proof or not depends of course on the meaning of the sentences involved in the proof. We may feel that a correct proof of the conjunction "*A* and *B*" must contain proofs of both *A* and *B*. But the only way to state this idea that I know is to appeal to a notion of canonical proof. We may demand that a canonical proof of a sentence "*A* and *B*" proceeds via canonical proofs of the two sentences *A* and *B*. Our intuitions about correct proofs of *e.g.* conjunctions may then be stated by saying that such proofs must be possible to transform to canonical proofs, *i.e.*, they must contain or consist of a method for finding a canonical proof. The general condition for asserting a sentence is thus that we possess either a canonical proof of the sentence or a method for finding one.

The statement of the rules for drawing consequences from an assertion meet a similar problem. The consequences of, *e.g.*, a conjunction "*A* and *B*" do not consist of just *A* and *B*. When the conditions for correct assertions are given, however, we have also a general condition for the drawing of consequences. To infer a conclusion is at the same time to assert the conclusion on the assumption of the premises, which can be correct only if we know a method for finding a canonical proof of the conclusion given (canonical) proofs of the premises. When the rules for drawing consequences are formulated in this way, we have already incorporated the requirement of harmony (discussed in Section 5) in the formulation itself.

The so-called reductions associated with the elimination rules in systems of natural deduction are nothing else than methods that verify the correctness of the elimination inferences in the above sense. As I have described elsewhere in more detail, Gentzen's Hauptsatz or my normalizations in natural deduction are just based on considerations of meaning of the kind described in Section 5. The meaning theoretic principles discussed here put the proof theoretic results mentioned into a more general context, and conversely, they get a concrete, substantial illustration by these results, which holds out some hope of the possibility of formulating a general pragmatic meaning theory along the lines discussed above.

7. A NON-REALISTIC PRINCIPLE OF TRUTH

As we have seen, the notion of truth is replaced by the condition for a correct assertion as constitutive element in a meaning theory of the kind discussed here. Nevertheless, it is of interest to see what status the notion of truth has in this theory, and it is not to be excluded that meaning is still determined by truth conditions.

In mathematics, where the condition for asserting a sentence is the possession of a proof of the sentence, we may say that the condition is to know the truth of the sentence. In general, the condition is of course less stringent, and it is enough to have good grounds for believing the sentence to be true. But it is clear from this that there is a way to derive the condition for asserting the sentence from the truth condition of the sentence (although there may be no such derivation from knowledge of truth conditions to knowledge of conditions for correct assertions) given that we know that the truth conditions in question are truth conditions.

However, if truth is just given by a number of equivalences corresponding to Tarski's adequacy condition possibly supplemented by the platonistic principle of truth, then there is nothing which legitimates this derivation of conditions for correct assertions from truth conditions. Hence, the classical theory of meaning and truth is certainly unable to explicate the conditions for correct assertions. It is rather the other way around: the notion of truth gets its fundamental characteristics from its relation to the conditions for correct assertions and is explicated in terms of these conditions.

This suggests that the notion of truth should be derivable from the conditions for correct assertions by defining a sentence to be true when there is some situation in which it can be correctly asserted. By "some situation" is here to be understood not an actual historical or future situation, *i.e.*, a situation that has been or will be realized, but a possible situation. In the case of mathematics, a sentence would be true if and only if there is some possible situation in which a proof of the sentence is constructed, *i.e.*, if a proof of the sentence can be given.

Outside mathematics, the conditions for correct assertions may allow assertion of sentences that in fact are not true. If a sentence is asserted in mathematics on the basis of what one thinks is a proof of it and it later turns out that the sentence is false, one would ordinarily say that the alleged proof was not a proof and that therefore the sentence was incorrectly asserted. But outside mathematics, one may want to say that a sentence was correctly asserted (on sufficient ground) although it later turned out that the sentence was false, *i.e.*, the grounds on which the sentence was asserted are still regarded as having been sufficient in the situation in question (although they are not so anymore). The definition of a non-realistic concept of truth suggested above, *i.e.*, the identification of truth with in-principle-possible-to-assert-correctly, is of course reasonable only if the conditions for correct assertion rule out the discussed possibility that a sentence is correctly asserted in one situation while its negation is correctly asserted in another situation.

Intuitionistic philosophers sometimes use true as synonymous with the truth as known, but this is clearly a strange and unfortunate use. We need a notion of truth where, without falling into absurdities, we may say, *e.g.*, that there are many truths that are not known today. But do we need a notion of truth that allows truths which are even in principle impossible to know?

The non-realistic concept of truth when at all reasonable agrees with the platonistic or realistic concept of truth in the case of sentences that are in principle decidable. Furthermore, the two concepts agree (in contrast to the intuitionistic one mentioned above) in allowing the existence of truths which in fact will never be known. What the above non-realistic principle or truth rules out is the existence of truths that are not even in principle possible to know.

The difference between the two principles boils down to this: on the platonistic principle, a truth condition for a sentence obtains or does not obtain independently of our means of recognizing that it obtains or fails to obtain, and we are then forced to admit that there may be truths that are in principle impossible to recognize (if we are not to assert unwarrantably that all problems are in principle solvable); on the non-realistic principle above, a truth is in principle always possible to recognize, but we must then refrain from asserting that a truth condition either obtains or does not obtain (again, in order not to assert that everything is solvable). Both principles respect the fact that we are not omniscient, but the platonistic principle does this by introducing ideas the need of which are not easily seen.

8. ATTEMPTS TO EVALUATE THE ISSUE

Although the general questions about truth and meaning raised by the dispute between classical and intuitionistic logic may be even more interesting than the dispute itself, we may try to evaluate the dispute in the light of our above discussions about meaning. The main question should be whether an adequate meaning theory can explain the meaning of the logical constants in such a way that the validity of the laws of classical logic is seen.

In the case of the law of the excluded third, a defence may be tried in two different ways. One possibility is to dismiss as too strong the intuitionistic condition for the correct assertion of a disjunction "*A* or *B*," *viz.*, the possession of a method for finding either a canonical proof of *A* or of *B*, and to explain the meaning of "*A* or *B*" by saying that the condition for its correct assertion is just the same as that for the sentence not-(not-*A* and not-*B*). In that case, the law "*A* or not-*A*" is just synonymous with the law not-(not-*A* and not-not-*A*). In other words, it becomes an instance of the law of contradiction and loses its position as an independent principle of logic. No objection can then be raised against it, but now it does not seem any longer to express what one thought it expressed.

The second possibility should be to find conditions for asserting "*A* or *B*" that ascribe some distinctive character to disjunction and make the law of the excluded third a valid principle independent of the law of contradiction. But this seems very unlikely, and in any case, one does not know any such conditions; the intuitionistic ones certainly do not give any support to this law.

When it comes to the question of how to support the law of double negation, there seems to be no choice of the same kind. There is no disagreement about the conditions for asserting a negation or an implication. That is, it seems impossible for the classical logician to find a stronger condition for the assertions of negations or implications than the one formulated by the intuitionist, in spite of the fact that the classical logician seems to feel that he understands negation and implication in a weaker sense than the intuitionist does. Furthermore, the rule of inferring the sentence A from the sentence not-not-A is not in general in harmony with the condition for asserting a double negation. Hence, it is difficult to see how an adequate meaning theory could be formulated in defence of the law of double negation.

Of course, these considerations do not show that we must immediately give up the deductive practice described by classical logic and adopt the one of intuitionistic logic. Our goal should be a reflective equilibrium between practice and theory. The difficulty of formulating a meaning theory which accords with a certain classical practice need not be seen as indicating that something is wrong with this practice, but as a weakness in our theoretical attempts so far to make this practice intelligible.

However, to stick indefinitely to a practice that we cannot make intelligible is a sign of irrationality. And the more we get convinced of having found correct theoretical principles, *e.g.*, about meaning, that do not agree with certain practices, the stronger must be the pressure to give up the conflicting practice.

REFERENCES

Dummett, Michael, "Truth," *Proceedings of the Aristotelian Society*, new series, vol. 59 (1959), pp. 141–162.

Dummett, Michael, "The philosophical basis of intuitionistic logic," in: *Logic Colloquim* '73, ed. by H.E. Rose *et al.*, Amsterdam, 1975, pp. 5–40.

Dummett, Michael, *Elements of Intuitionism*, Oxford, 1977.

Prawitz, Dag, "Ideas and results in proof theory," in: *Proceedings of the Second Scandinavian Logic Symposium*, ed. by J.E. Fenstad, Amsterdam, 1975, pp. 235–250.

Prawitz, Dag,"On the idea of a general proof theory," *Synthese* 27 (1974), pp. 63–77.

Prawitz, Dag, "Meanings and proofs: on the conflict between classical and intuitionistic logic," *Theoria* 43 (1977), pp. 1–40.

MICHAEL DUMMETT

(*University of Oxford*)

COMMENTS ON PROFESSOR PRAWITZ'S PAPER

I am in very great agreement with what Prawitz says. This is not surprising, since he is expounding ideas which, for some years, he and I have been developing, each influenced by the other in a way that makes it now impossible, even if there were any point in doing so, to disentangle who contributed what. I am grateful to him for the very kind acknowledgments to me that occur in the present paper, but have in turn to express my great debt to him in recent work concerning the application of these ideas to logic. I must also express my admiration for the concision and lucidity of the summary exposition which he gives in his paper of a range of ideas that it is not always simple to state or for others to grasp.

What is it to know the meaning of a sentence? It is, obviously, to grasp the role that that sentence has in discourse, in communication between speakers: that is, to know how, if occasion arises, to use that sentence oneself, and to understand its use by others. To say this much is in no way explanatory, since it conveys no clear idea of what it is to "use" a sentence, that is, of what a speaker does by uttering it. It is nevertheless worth saying, since it focusses attention on what it is that needs to be explained: the essential task of a theory of meaning is to display the conventional significance of an utterance of any sentence of the language, what difference, as it were, is made to the world by the fact that it has been uttered. Now, as Prawitz remarks, I have repeatedly pointed out that a set of equivalences of the form "*S* is true iff *P*," for every sentence *S* of the language, or a theory yielding such equivalences, cannot possibly be, simultaneously, an elucidation of the concept *true* and a specification of the meanings of the sentences of the language. If it is to serve as a theory of meaning, it will have to invoke some general principles, so far unstated, connecting the condition for the truth of a sentence with its use. Prawitz goes further: he says that, even if we were given these principles, the truth of a statement of the form "*S* is true iff *P*" cannot by itself determine the meaning of *S*; we should have to know that this equivalence stated *the* truth-condition of *S*, that is, as he puts it, that the right-hand side of the equivalence stated the proposition expressed by *S*. This argument is very

G.H. von Wright (ed.), Logic and Philosophy, 11–18.

terse. It assumes that those principles, whatever they may be, that are supposed to connect the truth-condition of a sentence with its use will not establish such a connection between use and just *any* condition which, as a matter of fact, is equivalent to the truth of the sentence. I should like to say, and should be interested to know if Prawitz agrees, that the connection can be only with that condition which represents the way in which the meaning of the sentence is *given* to us; that is to say, that condition which, in some sense that requires further elucidation, a speaker must implicitly know to be equivalent to the truth of the sentence if he is to grasp its meaning, and not to some condition whose equivalence to the truth of the sentence depends on facts of which a speaker may be quite unaware.

It is highly problematic whether, as I am suggesting, it is unobjectionable, let alone indispensable, to invoke the notion of knowledge in connection with meaning. Ought we to try to explain what endows the sentences of a language with the meaning that they have by simply describing what the speakers of the language actually *do* (including, of course, what they say), or ought we to explain it in terms of the *ability* of each speaker to speak the language, that is, to engage in a certain complex of activities involving the utterance of sentences? And, if the right strategy is to appeal to the conception of a speaker's linguistic ability, to what is often called his mastery of the language, ought this ability to be explained in terms of his having a body of *knowledge* of some kind? Prawitz bypasses these questions, which are very difficult to answer. Setting them aside, we are at this stage in the following position. If it can be said that the meaning of a sentence is given by the condition for it to be true, this can be justified only against the background of some general account of the connection between the truth-condition of an arbitrary sentence and its use, that is, the difference that is made to what happens by an utterance of it. Such an account cannot be expected to be correct when applied to every true statement of the form "*S* is true iff *P*": for each sentence *S*, there will be just one such equivalence to which the account may be correctly applied, and this equivalence will be that determined by the way in which the content of *S* is given in accordance with the internal structure of *S*. If it is possible to devise both (1) a detailed theory which will yield, for every sentence of the language, a corresponding truth-condition and (2) a general account of the connection between the truth-condition of any sentence and its use which, when applied to the truth-conditions yielded by the theory (1), will result in an adequate description of the practice of speaking the language, then we shall have what may be called a truth-conditional theory of meaning. Against the background of such a theory of meaning, we may correctly say that the meaning of a sentence is given by the condition for its truth, not indeed as meaning that only part (1) of the theory is needed, but as conveying that, in the context of the *general* account provided by part (2), we may say that the *specific content* of each sentence is determined by its truth-condition as derived from part (1).

But is this possible? Is a truth-conditional theory of meaning feasible? This is the question that most concerns Prawitz, and to which he is disposed to give a negative answer. Here, before even attempting a definite, if tentative, answer, some clarification of the question is of help. This is best done by considering a particular case, for instance the intuitionistic account of the meanings of mathematical statements. At first sight, we are inclined to say that this is plainly *not* a truth-conditional theory of meaning, on the ground that it may be stated without so much as mentioning the notion of truth, purely in terms of the notions of a mathematical construction and of such a construction's being a proof of a statement. On the other hand, an intuitionist is obliged neither to reject the concept of truth altogether nor even to deny that to grasp the meaning of a mathematical statement is to know the condition for it to be true. He interprets the truth of a mathematical statement as the existence of a construction which is a proof of it; and since, for him, to know what a mathematical statement means is to be able to recognise, of any construction, whether or not it is a proof of the statement, we may also legitimately say that it is to know what is required for that statement to be true.

This does not provide any ground for classifying the intuitionist theory of meaning as a truth-conditional one; what it shows, rather, is that we cannot characterise a truth-conditional meaning-theory as one for which it is correct to say that to grasp the meaning of a sentence is to know its truth-condition. We come nearer to what we intuitively want if we say that by a truth-conditional meaning-theory we understand one in which the notion of truth is itself *used* in specifying the particular content of the sentences. In the intuitionist theory of meaning, it is not used: the specific content of any sentence is fixed by stipulating what is to count as a proof of that sentence, and the notion of truth is then explained, in a manner uniform for all sentences, in terms of the notion of proof.

We may generalise this as follows. Any theory of meaning will be dissectable into two parts, in the same way as we already described a truth-conditional meaning-theory as being dissectable. Part (1) of the theory will give the meaning of the individual words of the language and the principles governing the formation of phrases and sentences, in such a way as to determine the specific content of every sentence of the language. In doing this, it will employ certain fundamental notions, for instance, that of denotation; these may be termed the *central notions* of the meaning-theory; they are, in effect, its theoretical concepts. Part (2) of the meaning-theory will establish the connection between the meaning of any sentence, as given in terms of these central notions, and its use; that is, between part

(1) of the theory and the actual practice of speaking the language; it will do this uniformly for all sentences, and so will not contribute to giving the *specific* content of any particular sentences. It may be that there exists a means of characterising the general notion of the truth of a sentence in terms of the central notions of the meaning-theory; and it may further be that the theory is such as to allow us correctly to say that to know the meaning (specific content) of any sentence is to know the condition for its truth. This will not, however, be enough to justify us in classifying the meaning-theory as a truth-conditional one. What is required for that is that truth should itself be one of the central notions of the theory, in the sense explained.

This characterisation of a truth-conditional meaning-theory only approximates to the principle of classification we desire; as it stands, it is too restrictive. To start with, a theory of meaning for a language containing indexical expressions will not employ a notion of the absolute truth of sentences, but of their truth relative to an occasion of utterance. Even more trivially, it is well known that a classical or realistic theory of meaning, based upon a classical two-valued semantics, can always be framed, in a Tarskian manner, in terms, not of the notion of the truth of a sentence, but of that of the satisfaction of a sentence by an infinite sequence of objects (elements of the domain). A classical meaning-theory so formulated perfectly fits the characterisation of a non-truth-conditional meaning-theory as I stated it: part (1) of the theory specifies the individual content of each sentence in terms of certain notions, including that of satisfaction but not including that of truth, and the notion of truth can then be defined uniformly for every sentence in terms of those central notions, *viz.*, as one that is satisfied by every sequence. To admit such a meaning-theory as non-truth-conditional would obviously render the notion completely nugatory, so we must reformulate our characterisation so as to exclude this case. This is of course easy enough to do if we can be sure that the examples cited are the only ones we need to worry about; but it is far from easy to be sure of that. How, for example, are we to treat Hintikka's game-theoretic semantics, a version of which will yield a classical logic? Here part (1) of the theory employs the notion of truth only for atomic sentences: the meanings of complex sentences are given in terms of a two-person game, a move in which is an utterance of a sentence of lower complexity than that uttered by the opponent, winning or losing the game then being defined in terms of the truth or falsity of an atomic sentence uttered. In terms of such a theory, the general notion of the truth of a sentence can be explained, uniformly, as the existence of a winning strategy for a game that begins with an utterance of that sentence; so, on the characterisation as stated, this should count as a non-truth-conditional meaning-theory. It is not immediately apparent how the characterisation should be framed so as to count such a theory as truth-conditional, nor, indeed, whether it is intuitively desirable to do so. The fact that a theory of meaning yields a classical logic is probably a necessary, but certainly not a sufficient, condition for classifying it as truth-conditional.

Leaving these difficulties unresolved, let us return to the question whether a meaning-theory of an outright truth-conditional form is feasible. In such a theory of meaning, every statement is assumed to have a determinate truth-value independently of our knowledge. If it is assumed that a speaker, in knowing the meaning of a sentence, knows the condition for it to be true, or, more generally, the condition for it to have any one of the truth-values one of which it must have, then we may argue as follows. Since, manifestly, not every meaningful sentence is such that we can get ourselves into a position in which we can recognize it as having any one of the possible truth-values, the knowledge of its truth-conditions cannot be exhaustively manifested by our capacity to recognise its truth-value. But since such knowledge cannot in general be taken to be explicit knowledge, on pain of a circular explanation of that in which the mastery of a language consists, such knowledge must, in some cases, be knowledge that cannot be fully manifested in any way, and is therefore without substance. We therefore have to restrict a speaker's knowledge to his grasp of those conditions under which each sentence can be recognised by us as having one or another particular truth-value; the attribution to him of further knowledge of what it is for it to have such a truth-value independently of our capacity to recognise it is spurious, and has no place in an account of the actual practical ability that constitutes mastery of a language.

Prawitz alludes to this form of argument, but does not endorse it. Its weakness lies in its resting on the assumption that we already set aside as highly problematic, namely that the institution of language has to be explained in terms of the ability of the individual speakers, and that this ability embodies knowledge of some kind. But he claims that the same conclusion may be reached without appealing to this assumption. The question is how part (2) of the truth-conditional meaning-theory can be constructed: that is, how it is possible to give a general method for deriving, from the specification of the truth-conditions of a sentence, the principles governing its actual use. If the sentence is an assertoric one, those principles are of two kinds: those governing when it may justifiably be asserted, and those governing what someone commits himself to by asserting it or accepting another's assertion of it. It may appear that there is no difficulty in deriving these principles from the truth-condition of the sentence: for it may seem obvious that, if someone knows what it is for the sentence to be true, he can deduce what counts as a ground for taking it to be true, and also what taking it to be true involves, that is, what it is to act on it. On reflection, however, this is not so obvious at all: how are we supposed to connect up the knowledge of a possibly transcendent truth-condition, one that we cannot directly recognise as obtaining, with our actual recognitional capacities? In any case, the advocate of a truth-conditional meaning-theory is in a dilemma. If he claims that a speaker actually knows the condition for the truth of each sentence, then he is open to the form of argument against the legitimacy of ascribing such knowledge to the speakers, or, rather, of taking it to be what constitutes their mastery of the language, that we earlier set aside. But if he denies that the speaker actually knows the truth-conditions of the sentences, it ceases to be in any way plausible that he has a means of deriving the actual use of those sentences from the specification of their truth-conditions. I think this is the purport of the very compressed argument given by Prawitz at the beginning of Section 6 of his paper; but I should welcome his comments.

As Prawitz says, if we become sceptical about the feasibility of a truth-conditional meaning-theory, the most obvious alternative is to take as the central notions of our theory of meaning, in terms of which part (1) of the theory will be framed, one or other of the two aspects of use themselves. One such choice is to take the specific content of each sentence as given by its assertibility-conditions; the other is to take it as given by the consequences of an utterance of it. Although, as contrasted with what happens in a truth-conditional meaning-theory, we are now taking as our central notions ones directly relating to actual linguistic practice, to the use of sentences, we shall still need a part (2) of our theory, just because there are two aspects of use: whichever we choose as providing the central notions of our theory, we shall need some uniform means of deriving the other. This is alternatively expressible by saying that the two aspects must be in harmony; and I am in agreement with everything Prawitz says on this question. Prawitz explains in outline the applications that he has made of this idea to logic. Here the assertibility-conditions of a complex statement can be regarded as embodied in the introduction rules for the various logical constants that appear in it. The consequences of making such a statement can then be taken as embodied in the elimination rules, and the process of derivation of the one from the other involves a procedure of justifying arbitrary forms of argument, including elimination rules, from the introduction rules taken as constitutive of the meanings of the logical constants. (This justification-procedure may be embodied in a definition of "valid argument," but that comes to the same thing.) As Prawitz remarks, the admission of elimination rules only if justifiable in this way guarantees harmony between them and the introduction rules, interpreted in terms of the reduction steps involved in normalisation. We have, therefore, a realisation of the idea adumbrated by Gentzen when he said that the introduction rules could be viewed as defining the logical constants and the elimination rules as consequences of those definitions.

I should like to make both a proviso and a comment. The proviso is this: we are not entitled to assume that everything that looks like a logical constant is a logical constant in a strict sense. If we are adopting a theory of meaning in terms of assertibility-conditions, then we must, for example, suppose that the meaning of "or" is given by the conditions for asserting a statement of the form "*A* or *B*," given the meanings of "*A*" and "*B*." But our theory of meaning, by itself, gives us no entitlement to assume that these conditions are statable by purely logical rules, that is, by ones which take no account of the internal structure of "*A*" and "*B*," or, at least, of their non-logical structure. In the context of such a meaning-theory, this yields a very strong condition for being a logical constant; but we have no right to assume that we know in advance which the logical constants are.

Now for the comment. Would it not be equally feasible to adopt the reverse procedure, and take the meaning of a sentence as given in terms of the consequences of uttering (asserting) it, that is, roughly, what accepting it would lead you to do that you would not otherwise have done? In the general case, this seems highly problematic, because of the often remarked fact that what difference a belief makes to your behaviour depends upon your wants. But possibly this is not an insuperable obstacle; after all, whether something is a ground for adopting a belief depends on your other beliefs, but, despite the amount of holism there is in the philosophical air at present, this does not make everyone despair of a meaning-theory in terms of assertibility-conditions. At any event, in the application of this idea to logic, there does not appear any intrinsic difficulty: it is a matter of taking the meanings of the logical constants as given by the elimination rules governing them, and formulating a procedure for justifying arbitrary forms of argument, including introduction rules, by reference to those elimination rules. There are, indeed, certain technical difficulties in the way of this programme; but I shall assume for the sake of argument that they can be overcome.

Now it would appear that, if this alternative approach is feasible, it will make no ultimate difference whether we adopt it or that favoured by Prawitz in terms of assertibility-conditions: there is supposed to be harmony between assertibility-conditions and consequences, and so it should not matter which we start from. But this view is superficial. Consider the application to logic. We now have, let us suppose, two alternative procedures for justifying forms of argument: what I shall call an *upwards* justification-procedure, which appeals to the introduction rules as giving the meanings of the logical constants, and the converse *downwards* justification-procedure, which in the same way appeals to the elimination rules. Now suppose that we start with a set of elimination rules: then the set of introduction rules that can be justified by reference to them (by a downwards procedure) is well defined. Now consider this set of introduction rules. The set of elimination rules that can be justified by reference to them (this time by an upwards procedure) is likewise well defined. The question is whether, by this means, we shall get back to the elimination rules we originally started with. If we do so, I shall say that that set of elimination rules was *stable*; in the same way, one can define the stability of a set of introduction rules. I emphasise that this property is not the same as that of harmony: harmony is, as Prawitz said, automatically guaranteed by the appeal to either type of justification-procedure.

As far as I can see, there is no general guarantee that any given set of elimination or of introduction rules will be stable in this sense. I think, however, that stability is a further intuitively reasonable condition to impose: that is to say, there are no intelligible meanings conferred on a set of logical constants by the stipulation of an unstable set of introduction or elimination rules. An alternative way of putting this, of course, is to say that if, for a given language, the set of introduction rules governing its logical constants is unstable, then no theory of meaning in terms of assertibility-conditions is feasible for that language; conversely, if its set of elimination rules is unstable, no meaning-theory in terms of the consequences of assertion is feasible. These remarks should, however, be taken in conjunction with the proviso I stated earlier: we should consider stability in relation to *all* the rules of inference governing complex statements, not only to those that are logical rules in the narrow sense.

RISTO HILPINEN

(*University of Turku*)

SOME EPISTEMOLOGICAL INTERPRETATIONS OF MODAL LOGIC

I

Our beliefs are *vague* in (at least) two ways: the propositions believed are vague (contain vague concepts), and the boundary between belief and unbelief is vague, that is, the concept of belief itself is vague. Vagueness of the latter type is illustrated by the following sentences:

(1) Light travels faster than sound.

(2) Large doses of vitamin C protect against the common cold and influenza.

(3) Olof Palme will be the prime minister of Sweden in 1980.

(4) David Hume was 5′ 6″ tall.

I believe (1) more firmly and definitely than (2): unlike (1), (2) is not completely free from doubt. I *think* that the social democrats will win the general elections in Sweden in 1979 and that Olof Palme will then become the prime minister again, but I am (at the present time, in 1978) considerably more uncertain about this than about the truth of (2). On the other hand, I have no opinion whatsoever about David Hume's height (except that he was neither a midget nor a giant), and am thus completely agnostic in regard to (4).

Here the boundary between belief and unbelief can be drawn in different ways. If we assume that belief implies complete absence of doubt, I can be said to believe (1), but not (2) or (3), whereas more liberal interpretations of "belief" may include both (1) and (2), and perhaps also (3), among my beliefs. Different ways of separating beliefs from unbeliefs correspond to different ways of making the concept of belief completely precise or to different *specifications* of the concept of belief.[1] In the above example every acceptable (or admissible) specification should include (1) in my belief set and exclude (4).

G.H. von Wright (ed.), Logic and Philosophy, 19–29.

Propositions (1)–(4) can also be regarded as representing different *degrees* of belief (or degrees of confidence): I assign the maximum degree of confidence to (1), various positive degrees to (2) and (3), and zero degree of confidence to (4). In the present context the expressions "degree of confidence" and "degree of belief" are not synonymous with "probability": zero probability does not correspond to zero degree of belief (*i.e.*, complete lack of opinion), but to extreme *disbelief*. Belief and *surprise* are correlative concepts: if a person believes that H, then he or she would be *surprised* by (the truth of) $\neg H$, but if he is completely agnostic regarding H, then neither H nor $\neg H$ would surprise him at all. In the example presented above, I would be extremely surprised if (1) turned out to be false — in fact, I cannot even conceive of the possibility that it might be false. I would be greatly surprised if future research were to disprove (2), only a little surprised by the falsity of (3), but neither (4) nor its negation would surprise me in the least.

II

G.L.S. Shackle and Isaac Levi have argued in several publications that the concept of surprise and the associated concept of degree of belief cannot be explicated in terms of subjective probability.[2] Shackle has developed a nonprobabilistic theory of surprise and belief in which a person's degree of belief in a proposition is defined as the surprise-value of its negation.[3] Shackle characterizes the concept of (degree of) surprise (the concept of *potential* surprise, as he calls it) in terms of nine axioms or postulates, five of which concern the formal properties of the concept of potential surprise.[4] These postulates are listed below under (5)–(9). (The notation used here is not in all respects identical with Shackle's.)

(5) Let $s(H)$ be the degree of potential surprise assigned to H. $0 \leqq s(H) \leqq z$.

For simplicity, I shall assume here that we can distinguish only a finite number of different degrees of potential surprise, and that these degrees are represented by nonnegative integers 0, 1, ..., z; thus s is a function which assigns some nonnegative integer i ($i \leqq z$) to every proposition H.

(6) "The degree of potential surprise associated with any hypothesis will be the least degree among all those appropriate to different mutually exclusive sets of hypotheses (each set considered as a whole) whose truth appears to the individual to imply the truth of this hypothesis."[5]

I shall simplify this axiom as follows: first, each of the "mutually exclusive sets of hypotheses" mentioned by Shackle will be regarded as a single proposition, and secondly, I shall represent the condition that the truth of a proposition G "appears to the individual [X] to imply the truth [of H]" by the strict conditional "$N(G \rightarrow H)$," where the concept of necessity may be thought of as being relative to X. Thus (6) can be expressed in the following simplified form:

(6*) If $H_1, \ldots, H_m$ is a set of mutually exclusive hypotheses ($N\neg(H_i \,\&\, H_j)$ if $i \neq j$) and $N((H_1 \vee \ldots \vee H_m) \leftrightarrow H)$, then $s(H) = \min\limits_j s(H_j)$ ($j = 1,2, \ldots, m$).

(7) All members of an exhaustive set of rival hypotheses can carry zero potential surprise.

By an "exhaustive set of rival hypotheses" Shackle means any set of mutually exclusive and jointly exhaustive hypotheses, *i.e.*, any set $\{H_1, H_2, \ldots, H_m\}$ such that $N(H_1 \vee \ldots \vee H_m)$ and $N\neg(H_i \,\&\, H_j)$ if $i \neq j$.

(8) When H is any hypothesis, the degree of potential surprise attached to the contradictory of H is equal to the smallest degree attached to any rival of H.

This condition follows from (6*). If $\{H, H'_1, \ldots, H'_m\}$ is an exhaustive set of rival hypotheses, $N((H'_1 \vee \ldots \vee H'_m) \leftrightarrow \neg H)$; thus (6*) implies that $s(\neg H) = \min\limits_j s(H'_j)$.

(9) At least one member of an exhaustive set of rival hypotheses must carry zero potential surprise.

In other words, if $\{H_1, \ldots, H_m\}$ is a set of mutually exclusive and jointly exhaustive hypotheses, $s(H_j) = 0$ for at least one H_j. The disjunction of m jointly exhaustive hypotheses is a necessary truth; thus (9) follows from (6*) and

(10) If H is necessary, $s(H) = 0$:

any necessary truth (or any statement necessary for X) carries zero degree of potential surprise. The assumption that both $s(H) = 0$ and $s(\neg H) = 0$ (or, more generally, that $s(H_j) = 0$ for every hypothesis H_j in some set of mutually exclusive and jointly exhaustive hypotheses) is consistent with Shackle's postulates (5)–(6) and (8)–(9); thus the system defined by these postulates satisfies condition (7). If the N-operator used in (6) and (10) is (re)interpreted as logical necessity, Shackle's system can be characterized in terms of the following three axioms:

(S1) $0 \leqq s(H) \leqq z$,

(S2) $s(G \vee H) = \min(s(G), s(H))$,

(S3) $s(H \vee \neg H) = 0$,

together with the rule

(SR) If $G \leftrightarrow H$ is a logical truth, $s(G) = s(H)$.

III

According to Shackle, degrees of potential surprise are degrees of possibility (or degrees of impossibility): zero potential surprise corresponds to *perfect possibility* and the maximum degree of surprise to impossibility or "absolute rejection" of a hypothesis. In fact, if we define, for each degree of potential surprise i, a possibility operator P_i by

(11) $P_i H =_{\mathrm{df}} s(H) \leqq i$,

Shackle's axioms (S1)–(S3) imply the familiar modal axioms

(P1) $P_i(G \vee H) \leftrightarrow P_i G \vee P_i H$

and

(P2) $P_i H \vee P_i \neg H$

for every i ($0 \leqq i \leqq z$). "H is possible" means here that H is possible in the opinion of some individual X; hence "$H \rightarrow P_i H$" is not valid. "$H \rightarrow P_i H$" holds only for $i = z$, but P_z is not a genuine concept of possibility at all: "$P_z H$" means that H is possible *or* impossible; thus "$P_z H$" (and consequently "$H \rightarrow P_z H$" as well) holds for every proposition H.

The degree of potential surprise associated with a hypothesis H can be regarded as a measure of *disbelief*; thus a person's degree of belief in H, $b(H)$, can be defined as the degree of potential surprise associated with $\neg H$:

(12) $b(H) = s(\neg H)$.

If we now define, for each i $(0 \leqq i \leqq z)$, a modal belief operator B_i,

(13) $\quad B_iH =_{\mathrm{df}} b(H) \geqq i,$

(11) and (12) imply

(14) $\quad B_iH \leftrightarrow \neg P_{i-1} \neg H \ (i \geqq 1)$

and

(15) $\quad B_0H \leftrightarrow P_0 \neg H \vee B_1H.$

Notice that B_0 is not a proper belief operator: "B_0H" is a tautology; it means that X either believes H to some degree or does not believe it at all. According to (P1)–(P2) and (14), the proper belief operators B_1, ..., B_z satisfy the standard axioms of the modal system D, *viz.*[6]

(B1) $\quad B_i(G \mathbin{\&} H) \leftrightarrow B_iG \mathbin{\&} B_iH$

and

(B2) $\quad B_iH \rightarrow \neg B_i \neg H.$

These principles may be termed the conjunction condition and the consistency condition for degrees of belief, respectively.

IV

Shackle's theory of surprise and belief can be regarded as a theory of the *vagueness* of belief, and as such it is clearly more adequate than the probabilistic theory of degrees of belief. According to this interpretation, different positive degrees of belief and the corresponding proper belief modalities B_1, ..., B_z represent various admissible specifications of the vague concept of belief. Degrees of belief in this sense must obviously range from lack of opinion (agnosticism) to complete confidence, not from extreme disbelief to absolute certainty.[7] As was pointed out above, all proper belief modalities B_i satisfy the standard principles of the modal logic of belief, in particular, they all satisfy the conjunction condition and the consistency condition: no matter how beliefs are distinguished from unbeliefs, believing a conjunction amounts to believing both conjuncts and conversely, and a person should believe H only if he or she does not believe its negation. If a vague belief-statement F is regarded as a *truth of* *the logic of belief* if and only if it is true for all admissible specifications of the concept of belief,[8] the logic of (degrees of) belief turns out to be a species of the standard modal logic, and there is no need to introduce many-valued "fuzzy" modal logics to account for the fact that epistemological possibility and belief are vague concepts.[9]

V

Isaac Levi has observed that "in ordinary language, degrees of belief are not measured along a single scale. We consider rather two scales: one measuring degrees of belief and one measuring degrees of disbelief."[10] However, if we regard degrees of belief as positive degrees of confidence and degrees of disbelief as negative degrees of confidence, the two scales mentioned by Levi (the b-scale and the s-scale) can be joined together into a single scale of degrees of confidence (d-scale) as follows:

(16) If $b(H) > 0$, $d(H) = b(H)$

(17) If $b(H) = 0$, $d(H) = -s(H)$.

Case (17) includes two possibilities: (i) If $b(H) = 0$ and $s(H) = 0$, $d(H) = 0$; this means that X is completely agnostic regarding H. (ii) If $b(H) = 0$ and $s(H) > 0$, $d(H) < 0$; thus a positive degree of surprise corresponds to a negative degree of confidence. The s-values and the b-values of various propositions can be recovered from their d-values as follows:

(18) If $d(H) > 0$, $b(H) = d(H)$ and $s(H) = 0$.

(19) If $d(H) = 0$, $b(H) = 0$ and $s(H) = 0$.

(20) If $d(H) < 0$, $b(H) = 0$ and $s(H) = -d(H)$.

According to (S1)–(S3) and (16)–(17), the d-function satisfies the following conditions:

(21) $-z \leqq d(H) \leqq +z$

(22) $d(H) = -d(\neg H)$

(23) $d(G \vee H) \geqq \max(d(G), d(H))$

(24) $d(G \mathbin{\&} H) \leqq \min(d(G), d(H))$.

The degree of confidence assigned to a conjunctive or a disjunctive proposition is not always determined by the degrees assigned to the conjuncts or the disjuncts. However, if $d(G) > 0$ and $d(H) > 0$, then $d(G \,\&\, H) = \min(d(G), d(H))$ (in accordance with the conjunction condition for degrees of belief), and if $d(G) < 0$ and $d(H) < 0$, then $d(G \vee H) = \max(d(G), d(H))$ (in accordance with the disjunction condition for degrees of disbelief).

VI

The possible worlds semantics of modal logic can be applied to the Shacklean theory of degrees of possibility and belief in a straightforward manner. Let $W = \{u, v, \ldots\}$ be a set of possible worlds and let U_0 be a subset of W which contains all *perfectly possible* or *perfectly plausible* worlds. A (precise) statement H is interpreted as asserting that the actual world belongs to a certain subset of W, $|H|$ or the *truth set* of H: H is true in every $u \in |H|$ and false in all other possible worlds. A proposition H is perfectly possible if and only if H is true in some perfectly plausible world, and H is believed by X (to some positive degree) if and only if H is true in all perfectly plausible worlds. Various specifications of the concepts of possibility and belief can be defined in terms of David Lewis's *systems of nested spheres*: A set of sets of possible worlds $\mathcal{N}_{u,X} = \{U_0, U_1, \ldots, U_y\}$, where each $U_i \subseteq W$, is a (finite) *system of nested spheres around* U_0 if and only if (i) U_0 is a subset of every $U_i \in \mathcal{N}_{u,X}$, and (ii) $\mathcal{N}_{u,X}$ is nested, that is, for every U_i and U_j in $\mathcal{N}_{u,X}$, $U_i \subseteq U_j$ or $U_j \subseteq U_i$.[11] The system $\mathcal{N}_{u,X}$ represents the opinions of some person X in some possible world u; hence the subscripts "u" and "X." If $y = z - 1$, the degrees of possibility $P_0, P_1, \ldots, P_{z-1}$ and the proper belief modalities $B_1, B_2, \ldots, B_z$ can be defined as follows:

(25) P_iH is true (for X in u) if and only if $|H| \cap U_i$ is nonempty ($i = 0, 1, \ldots, z-1$).

(26) B_iH is true (for X in u) if and only if $U_{i-1} \subseteq |H|$ ($i = 1, 2, \ldots, z$).

Different spheres U_i can be taken to represent different *distances* from the innermost sphere U_0, and the degree of possibility of a given proposition H can thus be intuitively understood as its distance or deviation from perfect possibility.

VII

Degrees of possibility can also be interpreted as degrees of support, if the concept of support is thought of as being defined in terms of eliminative induction. According to the elimination theory of induction, a hypothesis H is supported or confirmed by a test or a sequence of tests F if and only if (i) H is consistent with F, *i.e.*, F does not refute H, but (ii) some alternative hypothesis H' (some rival of H) is refuted by F. Let us now assume that the spheres $U_i \in \mathcal{N}_{u,X}$ are introduced as follows: Let $E = \langle E_0, E_1, \ldots, E_y \rangle$ ($y = z-1$) be a sequence of mutually independent tests (or *test-statements*), and let CE_j be the conjunction of the first $j+1$ tests in the sequence E. Let U_i ($i = 0, 1, \ldots, y$) be the truth-set of CE_{y-i}; thus

$$
\begin{aligned}
(27) \qquad U_y &= |CE_0| = |E_0|, \\
U_{y-1} &= |CE_1| = |E_0| \cap |E_1|, \\
&\vdots \\
U_0 &= |CE_y| = |E_0| \cap \ldots . \cap |E_y|.
\end{aligned}
$$

Each test-statement CE_j in the sequence $CE = \langle CE_0, CE_1, \ldots, CE_y \rangle$ is stronger or more stringent than its predecessor CE_{j-1} in the sense that any hypothesis H not refuted by CE_j has not been refuted by its predecessor CE_{j-1}, any H refuted by CE_{j-1} is also refuted by CE_j, and some H not refuted by CE_{j-1} may be refuted by CE_j. The degrees of possibility defined by (25) can now be interpreted as degrees of support. According to this interpretation, H is perfectly possible if and only if it is consistent with CE_y, and hence with all tests CE_j; in this case the test sequence gives perfect or maximum support to H. "$P_z H$" allows the possibility that even the weakest test CE_j, CE_0, fails to support H (falsifies H); this corresponds to zero degree of support. Other degrees of support correspond to intermediate degrees of possibility: in general, we may read "$P_i H$" as "H" is supported (by CE) to the degree $z-i$."[12]

VIII

The analysis of inductive support outlined above resembles that presented by L. Jonathan Cohen in his book *The Implications of Induction*.[13] Cohen also defines degrees (or grades) of support in terms of a sequence of increasingly stringent tests or experiments, and assumes that the degree of support of H increases with the strength of the tests that fail to falsify H. According to Cohen, the tests CE_j involve a number of experimental factors or variables v_i, ordered according to their importance and relevance to the hypotheses under consideration. The hypothesis can be tested under various circumstances by manipulating the variables v_i. In the weakest test (CE_0), the experimental variation is at its minimum: no relevant variables are manipulated at all, but they all are assumed to have some constant or "normal" value. The second test involves all possible variants (values) of the first variable (v_1), the third test contains trials in all combinations of the variants of v_1 and v_2, and so on.[14] The degree of support H depends on the complexity of the tests that fail to refute H. Cohen represents various degrees of support in terms of a family of indexed modal operators M_i,

$$M_0, M_1, \ldots, M_e, M_{e+1}, \ldots, M_d.$$

However, he assumes that the operators M_i correspond to various degrees of necessity as well as support: according to Cohen, $M_0, \ldots, M_e$ represent increasing degrees of inductive support from zero support to "full" inductive support or "the level of establishment," and the remaining operators $M_{e+1}, \ldots, M_d$ correspond to degrees of necessity; "M_dH" means that H is logically true.[15] Cohen argues that all positive degrees of support satisfy the conjunction condition

(28) $\quad M_i(G \,\&\, H) \leftrightarrow M_iG \,\&\, M_iH$

and the consistency condition

(29) $\quad M_iH \rightarrow \neg M_i \neg H;$

thus Cohen's support operators resemble the concept of necessity (or the concept of belief) rather than the concept of epistemic possibility. But these principles are not warranted by Cohen's characterization of the concept of inductive support in terms of the "capacity [of a hypothesis] to pass cumulatively tougher and tougher tests, *i.e.* to resist falsification by cumulatively richer and richer combinations of relevant factors":[16] a hypothesis is capable of passing a test if and only if it is consistent or *compossible* with the results of the test (or possible *given* the test-results). If the concept of support is defined along the lines proposed by Cohen, its logic should resemble the logic of possibility, not that of necessity.

According to the analysis of support presented in Section 7, a test sequence CE will give a positive degree of support even to a hypothesis inconsistent with CE, if some tests CE_j included in the sequence fail to refute the hypothesis. Cohen's theory of support has this consequence too, and it has been criticized on this point,[17] but Cohen has defended his theory as follows:

The grade of inductive support attributed to H represents the inability of certain complexes of relevant factors to cause H's falsification. If the introduction of a further factor does cause H's falsification, that should not be allowed to give H zero-grade support. For then H would have been put on the same level as a hypothesis that could not stand up to any test at all. H would not have been given due credit for the extent to which it is, in fact, reliable: it would not have been given due credit for the various combinations of circumstances in which it holds good.[18]

In other words, even if CE is conclusive evidence that a hypothesis H is false, it may show that H possesses some degree of reliability, *i.e.* that H implies true predictions in some combinations of circumstances. In such cases CE will give H a positive degree of support: even a false hypothesis should be given some credit if it is not *trivially* false, that is, if the proof of its falsity requires sophisticated tests. In this respect Cohen's concept of support is akin to Popper's concept of verisimilitude (truthlikeness), and this also suggests that it is (logically) analogous to the concept of possibility, not the concept of necessity. The concept of verisimilitude does not satisfy the conjunction principle or the consistency principle: two mutually incompatible hypotheses or theories may both possess a high degree of verisimilitude.[19]

The conjunction principle and the consistency principle may be acceptable in certain special cases, *e.g.*, when the tests CE_j concern only a simple universal generalization $G=(x)(Rx\rightarrow Sx)$ and its "modifications" of the form $(x)(Rx\ \&\ C_i x\rightarrow Sx)$, where C_i is some combination of the values of various "relevant variables" v_i,[20] but they cannot be regarded as general principles of the logic of support if the concept of support is defined in terms of eliminative induction. According to (28) and (29), two conflicting hypotheses H and H' cannot both be supported by the same data. If a hypothesis H is supported to any degree $i>0$, its rivals must necessarily have zero degree of support. This means that if a set $\mathscr{H}$ of alternative hypotheses is tested in terms of the experiments CE_i, even the weakest test in the series, CE_0, must be strong enough to eliminate all hypotheses $H\in\mathscr{H}$ except one. According to Cohen's logic of support, a hypothesis is supported by certain data only if the data rule out all alternative hypotheses. This conception of inductive support is obviously inconsistent with the traditional view of eliminative induction as a process in which the degree of support of a hypothesis is gradually increased as its rivals are falsified by increasing empirical evidence. According to the traditional view, a hypothesis is supported (or confirmed) by certain data if *some* alternative hypothesis is falsified by the data. Cohen's confusion on this point is apparently due to the mistaken assumption that the degrees of support and degrees of necessity form a logically homogeneous scale of modal categories: he assumes that the whole scale satisfies conditions which in fact can be plausibly ascribed only to a certain part of it.

NOTES

1. The expression "specification" is used in this sense by Kit Fine in "Vagueness, Truth, and Logic," *Synthese* 30 (1975), pp. 265–300 (p. 268).

2. See G.L.S. Shackle, *Expectation in Economics*, Cambridge University Press, Cambridge 1952, and *Decision, Order and Time*, Cambridge University Press, Cambridge 1961, Part II; Isaac Levi, "On Potential Surprise," *Ratio* 8 (1966), pp. 107–129, *Gambling With Truth*, Alfred A. Knopf, New York 1967, pp. 135–138, and "Potential Surprise in the Context of Inquiry," in: C.F. Carter and J.L. Ford, eds., *Uncertainty and Expectations in Economics*, Basil Blackwell, Oxford 1972, pp. 213–236.

3. This definition has also been proposed by H.H. Price in *Belief*, George Allen & Unwin, London 1969, p. 276.

4. *Decision, Order and Time*, pp. 80–81.

5. *Ibid.*, p. 80.

6. *Cf.* E.J. Lemmon and Dana Scott, *An Introduction to Modal Logic*, American Philosophical Quarterly Monograph No. 11 (ed. Krister Segerberg), Basil Blackwell, Oxford 1977, pp. 50–52. The affinity between Shackle's system and modal logic has been pointed out by C.L. Hamblin in "The Modal 'Probably',". *Mind* 68 (1959), pp. 234–240.

7. Many philosophers and linguists have attempted to explicate vagueness in terms of *degrees of truth.* According to this approach, "X believes H to the degree i" means that the degree of truth of the statement "X believes that H" is i (or i/z).

8. This definition is due to Kit Fine, "Vagueness, Truth and Logic": "A vague sentence is true if it is true for all admissible and complete specifications" (p. 278).

9. Something like this has been suggested by George Lakoff, "Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts," *The Journal of Philosophical Logic* 2 (1973), pp. 458–508 (p. 467).

10. "Potential Surprise in the Context of Inquiry," p. 235.

11. See David Lewis, *Counterfactuals*, Basil Blackwell, Oxford 1973, pp. 13–14 and 97–98. Notice that the expression "... around U_0" used here does not mean the same as David Lewis's expression "... around i."

12. This may seem an abnormally weak notion of inductive support: to say that certain data support a hypothesis does not mean *only* that the hypothesis is consistent with the data. The present interpretation of "P_i" reflects only certain simple formal features of the concept of support (as defined by the elimination theory). But perhaps it indicates also the insufficiency of the purely eliminative view of induction.

13. Methuen & Co., London 1970. See also L. Jonathan Cohen, *The Probable and the Provable*, Clarendon Press, Oxford 1977, chapters 13 and 14.

14. *The Implications of Induction*, p. 53. For example, if the tests concern the hypothesis "all ravens are black," the relevant variables might include the geographical location and the age of ravens. *Cf.* also Cohen, *The Probable and the Provable*, pp. 129–140.

15. *The Implications of Induction*, pp. 210–218.

16. L. Jonathan Cohen, "What has Inductive Logic to do with Causality?" (forthcoming in the Proceedings of a Conference on the Applications of Inductive Logic, Oxford 1978, Clarendon Press, Oxford).

17. *E.g.*, by Isaac Levi in "Potential Surprise: Its Role in Inference and Decision Making" (forthcoming in the Proceedings of a Conference on the Applications of Inductive Logic, Oxford 1978).

18. *The Probable and the Provable*, p. 135.

19. For the logic of verisimilitude, see Risto Hilpinen, "Approximate Truth and Truthlikeness," in Marian Przełęcki, Klemens Szaniawski and Ryszard Wojcicki, eds., *Formal Methods in the Methodology of Empirical Sciences*, D. Reidel Publishing Company, Dordrecht 1976, pp. 19–42.

20. *Cf. The Probable and the Provable*, pp. 182–183. In his defense of (28) and (29) Cohen discusses mainly cases of this kind; see also *The Implications of Induction*, pp. 64–65. If H is a "modification" of G (in Cohen's sense), it is a logical consequence of G. If the set $\mathscr{H}$ of "relevant" hypotheses consists of some universal generalization G (consistent with a given test CE_j) and its modifications, $\mathscr{H}$ is obviously logically consistent and satisfies the conjunction principle.

Ruth Barcan-Marcus

(*Yale University*)

HILPINEN'S INTERPRETATIONS OF MODAL LOGIC

A task of epistemology has been to sort out and analyze an intricate network of terms and related concepts. Of that plethora of terms and concepts, some seem more clearly descriptive of mental or psychological events in the lives of epistemological subjects. We say of ourselves and others that on some particular occasion we imagined that *p*, or decided that *p*, or entertained *p*, or came to accept *p*, or came to believe that *p* where *p* is some proposition. Of those terms that characterize mental or psychological events, some seem to be descriptive of cognitive or more appropriately "doxastic" feelings which a subject might have with various degrees of intensity when entertaining a certain proposition under suitable circumstances. A person may feel surprise or feel doubt or feel confidence that a certain proposition is true. Whether those psychological events are appropriately described as feelings has been questioned, particularly in the case of doubt or confidence. Still it is not unusual to claim that we feel doubt or confidence. Surprise is less controversial. Witness "He nearly fainted with surprise."

There is another class of doxastic terms which although mental or psychological seem more descriptive of dispositions of subjects rather than being descriptive of specific events in their lives. When we say of a person that he believes or accepts or doubts or has confidence in some proposition, we do not mean that he is in the perpetual thrall of doxastic feeling. An analysis of such terms might involve some claims about what a person would feel under certain hypothetical conditions, but it would also involve many claims about what he would think or do or say under certain hypothetical conditions. If a person believes *p*, then it is plausible to believe that in the hypothetical circumstance of his learning that *p* was false, some feeling of surprise would be evoked. Under certain hypothetical circumstances, if *p* is believed, then entertaining *p* might evoke feelings of confidence. But there are many other grounds which support the claim that someone believes *p*. If he is minimally rational, he will not believe *p* and believe not-*p* at a given time. He will take *p* into account as a premise in coming to conclusions in making plans for action, and so on. An

G.H. von Wright (ed.), Logic and Philosophy, 31–36.

adequate analysis of belief is, as we know, a complicated affair.

Furthermore, a general theory of belief is made even more complex by the fact that specifying the way a person's beliefs figure in his decisions and actions where we might stretch "action" to include coming to a conclusion from premises, depends on other parameters the values of which may vary between individuals. The extent to which belief in a set of propositions carries over to logical or mathematical consequences of that set depends on the deductive or mathematical abilities of the subject. How a person's beliefs will enter into his plans for action will depend on his desires and the extent to which his desires impede his rationality. It will also depend on his courage, his inertia and so on. If we also allow, as seems plausible, that given someone believes something, his belief has a degree attached ranging from, in the limits, non-belief to certainty and if the dispositional analysis must now include claims about how someone would act under some degree of belief in p, then there is the problem of how certain parameters, temperamental if you like, vary from case to case. How strong must the belief be? Is the person cautious, impulsive; does he take risks?

In addition to such more purely doxastic notions such as surprise and belief there are those epistemological terms which link doxastic notions to facts about the world. "Knowing that" is such a concept. Leaving aside for the moment whether an analysis of "knows that" as justified true belief is sufficient, it is generally accepted that for "x knows that P" to be true, P must be true which precludes a wholly dispositional account of "knows that." The truth of a knowledge claim, unlike the truth of a belief claim about P requires that P be true. Furthermore, if justification also figures in the analysis, it is required that where P is the kind of proposition which warrants support (unlike for example claims about being in pain or experiencing a red patch in a visual field) the reasons and chains of *argument given* in justification of P do in fact support it. Whether certain reasons or chains of argument do support P is presumably a matter independent of anyone's purely doxastic feelings, dispositions or attitudes. Indeed it is that separation which permits us to characterize rationality in an epistemological subject in terms of the degree to which he gives as reasons in justification of an hypothesis, only propositions and chains of inference which support it in accordance with correct canons. If a subject is wholly rational as well as omniscient, the order and connection of his beliefs will reflect the order and connection of things.

A complete account of the canons which validate inferences has yet to be given. It is the subject matter of theories of induction and deduction; of metamathematics and theories of statistical inference. The extent to which such theories are interreducible also remains open and debated. Still we have considerable grasp of those canons of *evidence and inference* which permit us to evaluate to some considerable degree the extent to which an hypothesis is supported by given premises and chains of inference. The last section of Hilpinen's paper is a further contribution to that subject and is more properly seen as a modal interpretation of inductive, rather than epistemic, logic.

Before proceeding to comments on Hilpinen's paper, it is important to mention a third and troublesome category of terms which are not unambiguously doxastic like "believes that" or clearly mixed like "knows that" but are rather equivocal as between doxastic and non-doxastic uses. We have already mentioned the philosophers' use of "conceivable" as "logically possible." "Is probable," "is acceptable," "is necessary," "is certain" are further examples. The equivocations are not accidental since the variant uses may share many logical and semantical features. It is those shared features which permit the adaptation of modal logic and semantics to a variety of interpretations of the modalities. The latter part of Hilpinen's paper may be seen as an application of David Lewis's semantics of possible worlds with its theory of comparative similarity of worlds to two recalcitrant notions: the doxastic notion of degrees of belief and the inductive notion of degrees of support.

My first group of comments are concerned with Hilpinen's adaptation of Shackle's theory as a basis for analysis of belief and degrees of belief. Shackle's is a theory about decision making under conditions of uncertainty. What in an agent is the interplay of belief and desire which leads to an outcome where that outcome is uncertain? Shackle was particularly concerned with economic decision making which could be tested in the market place. If we take outcomes to be propositions, Shackle makes the reasonable assumption that there are three possible belief stances with respect to an outcome or more generally a proposition. It can be believed, its negation can be believed or it can be doxastically neutral where neither it nor its negation is believed. That a fair coin will fall one way may have a probability of 1/2, but we are doxastically neutral to the outcome. Shackle takes the further step of defining those stances in terms of surprise.

The underlying primitive concept in Shackle's theory is one which corresponds to a doxastic feeling: the feeling of surprise. It can be characterized by supposing that for a given subject, any proposition has a surprise value, a degree of potential surprise, with the range of values from zero for not surprising to some maximum value. The surprise value of a proposition for a subject is a measure of the degree of surprise it would engender in that subject if it were true. The belief value of a proposition is the surprise value of its contradictory. Shackle's axioms which specify how surprise values are assigned have been neatly represented by Hilpinen, but his "simplification" of one of them is clearly questionable. Shackle saw his theory as empirical on the basis of which behaviour under conditions of uncertainty can be predicted. If it is an empirical theory, then Axiom SR is surely unjustified. It claims that if two propositions are logically equivalent, then their surprise values are identical and that is clearly false. We might very well, and often do, disbelieve propositions which, as it turns out, are entailed by strongly believed premises. The criticism cannot be directed against Shackle, who requires only that "it appear to the individual" that the entailment holds. But if Shackle's more constrained axiom is assumed instead, there would be no easy application of modal semantics as Hilpinen proposes. Hilpinen's simplification is a shift from the doxastic "appears to x to imply" to the modal strict implication.

Deeper difficulties arise from the definitions. The axioms are concerned only with the assignment of surprise values. But the presuppositions and definitions presume to define, in terms of potential surprise, "belief," "degree of belief," "degree of possibility" (in the sense of conceivability). "Degree of confidence" is also definable. It would appear that for Shackle, such notions and their cognates are all to be understood wholly in terms of a single doxastic feeling. But as already indicated, feelings, even potential ones, are an insufficient and dubious basis for a theory of belief. It hardly stretches the imagination to conceive of someone with a very sluggish affective life who doesn't feel surprise at all but may still be correctly described as having beliefs. He may even rank his beliefs, but those rankings may have nothing to do with some ranking of affective intensity.

Perhaps potential feelings such as surprise at a proposition's being true are in normal cases a fairly reliable symptom of someone's having a belief in its contradiction and the degree to which he has it, just as a pain of a certain kind on the left side and its intensity is a reliable symptom of someone's having appendicitis and the severity with which he has it, but a complete account of the disease is not reducible to one of its potential symptoms.

Still as an empirical theory, Shackle's is suggestive. It suggests an opportunity for cognitive psychologists to conduct simple experiments. Given a range of surprise values with the explanation that 0 is for no surprise at all and the maximum value z is for total surprise, would a proper sample of subjects in fact make the belief assignments and confidence assignments postulated and derived by the theory? My suspicion is that Shackle would be in for a surprise. I suspect for example that despite Shackle's vigorous theoretical opposition, probability estimates will figure in the assignment of confidence values. According to Hilpinen's elaboration of Shackle's theory, if the degree of confidence in p is greater than zero, then it equals the belief value of p. But it seems reasonable to conjecture that someone would assign a positive degree of confidence to a fair coin falling heads although he might also fail to be surprised if it did not fall heads, and fail to be surprised if it did.

I also suspect that being asked to assign surprise values to certain kinds of propositions would produce only puzzlement. Would you be *surprised* if $2+2$ were not equal to 4, or if some instance of excluded middle were false? I am also reasonably sure that the principles, such as the one which assigns as a belief value to the conjunction of two propositions the minimum belief value of its conjuncts, must be conditional, particularly with respect to propositions which describe uncertain outcomes. As a critic of Shackle has pointed out, I might believe to some degree that a child will have one blue eye, and also believe perhaps more strongly given some knowledge of genetics and the coloration of the parents' eyes that it will have one brown eye. Given that there are cases of variant color of eyes in a single person, the propositions are not even exclusive. But my belief in the child's having one blue eye and one brown one is far less than either taken alone.

I should like to turn now to some comments on Hilpinen's application of Lewis's modal semantics to an account of degrees of belief. It is an interesting project and in the least, of heuristic value, but it carries with it a large burden inherited from Lewis. Lewis asks us to take as true that there are indefinitely many possible worlds. Among them there are those which are similar to a given world and to a greater or lesser degree. Furthermore, he supposes that we can group similar worlds into sets where all of the members of each set are within a certain degree of similarity to a given world, this one for example. The structure gives rise to a theory of comparative possibility in terms of comparative similarity of worlds. Relative to a given world u, a proposition which is true only in a world with low similarity to u will have low possibility relative to u. It is difficult to classify the kind of modalities with which Lewis is dealing. They are not logical modalities. Metaphysical modalities have been suggested. In any case we are now familiar with the difficulties of Lewis's view. We can discount Lewis's astonishing realism with respect to possible worlds and his belief that possible worlds do not differ in kind from the actual one. It is not, I believe, crucial to the analysis. What is more crucial is the elusiveness of the central notion of similarity, as well as some of its odd consequences. A world like this one, except for one displaced nail, would seem very similar. But as we know, for want of a nail a shoe might be lost, for want of a shoe a horse might be lost, and mixing sources, for want of a horse a kingdom might be lost. Furthermore, if worlds are to be compared by comparing propositions true in each, among those propositions are modal ones and there is a clear circularity. There are further difficulties about allowable but questionable inferences and apparently true counterfactuals that turn out to be false which we do not have time to elaborate. Not all of the difficulties carry over to a doxastic interpretation. In a way, a doxastic reading, appropriately construed, gives the structure more plausibility. If the spheres are nested in accordance with what a subject *believes* to be degrees of similarity, then no metaphysical court sits in judgement. There is of course the question as to whether similarity judgements can always plausibly be made. Hilpinen has told us very little about how the nested spheres are ordered on the doxastic interpretation, and because of the brevity of his paper I am somewhat uncertain as to the details of the proposed ordering. In any case, the adaptation of Lewis's semantics to doxastic notions is remote from Shackle's original empirical theory. It is devoid of feeling and replete with worlds.

The use of Lewis's modal theories for an analysis of the concept of inductive support is more an adaptation of some formal features of his analysis rather than an application of the full-blown semantics of possible worlds. It is a fruitful application and close, as Hilpinen has noted, to Jonathan Cohen's modal analysis in *The Implications of Induction.* It is not, however, an epistemological application for it concerns relationships between propositions — those which are hypotheses and those which are test statements — and says nothing about knowledge, belief or cognate doxastic and epistemological notions. A theory of inductive support will of course figure in an account of knowledge and rational belief. Many of the problems which plague Lewis's fullblown theory are absent in this more modest application. It is entirely plausible to suppose that tests can be ordered and compared with respect to stringency in a far more coherent way than worlds can be ordered with respect to some elusive similarity relation.

As for Hilpinen's criticism of Cohen's modal analysis of inductive support, I believe he is correct in faulting Cohen for placing degrees of possibility in the sense of inductive support as the initial segment of a scale of degrees of necessity, but it is arguable. Cohen's defense of one of the consequences of the analysis, that a test sequence can lend support, to a degree, to a failed hypothesis, also seems correct. As all scientists know, the juncture at which an hypothesis fails, and the tests which it has survived, is informative if not publishable.

WOLFGANG STEGMÜLLER

(*University of Munich*)

TWO SUCCESSOR CONCEPTS TO THE NOTION OF STATISTICAL EXPLANATION

I. INTRODUCTION

There are two mistakes often committed in modern philosophy of science. I became aware of them only a short time ago because I myself committed them repeatedly. First, most attempts to explicate a metascientific term disregard in a particular way the distinction between *special* philosophy of science and *general* philosophy of science. The first one has to do with particular theories; the second one abstracts from particular theories. There is a large number of philosophers of science who believe that we should try to explicate all important metascientific notions, like "theory," "disposition," "law," "confirmation" or "corroboration" and, of course, "explanation," on the abstract level of general philosophy of science. This attitude mirrors the conviction that such explications are possible.

I no longer belong to this group of philosophers. Within general philosophy of science the most one can do, in my opinion, is to work out a general frame and to formulate some necessary conditions. Only in a negative case, when no explication is possible, may one get a definite result at a general level. *I shall try to present such a negative result in this paper.*

The second error consists in an overestimation of logical, in particular of semantic, methods. In many cases, the *pragmatic* circumstances are of the greatest importance as well. *To this point*, too, *I shall try to give an illustration.*

As far as the term "*explanation*" is concerned, it seems to me that it designates a large family of notions which divides into three main subfamilies. The items of the *first family* are usually called "*deductive-nomological explanations of facts.*" Nobody seems to have been able to give necessary *and sufficient* conditions for these concepts, presumably for the reasons just mentioned. But there is general agreement that the explanandum sentence of such an explanation must be *true.*

The "*explanations of laws and theories*" belong to a second family. Here, in the overwhelming majority of cases, we will have to do not with strict, but with approximative explanation. Thus, explanation will usually amount to approximative reduction.

G.H. von Wright (ed.), Logic and Philosophy, 37–52.

The third family contains the so called "*statistical explanation.*" There is no general agreement about their most basic structures. Hempel, *e.g.*, tried to parallel them with the "deductive-nomological" case, replacing the deterministic laws of the explanans by suitable statistical ones and weakening the deductive argument to an inductive one. Salmon, by contrast, tries to show in his "relevance theory" that statistical explanations are not arguments at all.

I shall start my discussion with the following:

Basic dilemma (A) consisting of two parts:

(1) Various authors have given us clear examples of simple statistical explanations.

(2) No explicate of "statistical explanation" satisfies simultaneously the following three minimal requirements of adequacy (which are so trivial that, as far as I know, nobody has ever formulated them explicitly):

(2.1) Those events for which explanations are given must not only be *possible* states of affairs; rather they must be *facts* (formally speaking: the explanandum sentences describing them must be *true*).

(2.2) If somebody claims to be able to give an explanation, then he must be in a position to tell *what* his or her alleged explanation explains. To this second requirement the following one can be added:

(2.3) Every explanation must be reconstructible as an answer to a why-question.

I shall concentrate mainly on the second part (2) and end with the suggestion that the term "statistical explanation" ought to be given up and replaced by *two successor concepts* of a very different structure.

As far as part (1) and the dilemma (A) is concerned, I shall restrict myself to very few remarks. My justification for this consists in the simple statement that overcoming the dilemma (A) is not a task for the philosophy of science but is rather the business of a psychology of the use of metascientific terms, like "statistical explanation."

II. STATISTICAL SINGLE CASE SUBSTANTIATION

The cases to be treated here correspond to those of deterministic predictions, where the occurrence of the predicted events can logically be deduced from the explanans, but where the question remains completely open whether the *reasons* given in the explanans may or may not be interpreted as *causes* of the occurring events. It is well known that Hempel tried to parallel statistical explanations as far as possible with the deductive-nomological type, the two main differences being that the deterministic laws are replaced in the explanans by statistical laws and that the deductive argument is transformed into an inductive one. It seems to me that the best improvement of Hempel's approach is due to Gärdenfors. I shall therefore speak of the Hempel-Gärdenfors line of thought. Since statistical single case substantiations *are arguments*, our present topic belongs to this Hempel-Gärdenfors line.

I am aware of the fact that various details of the following analysis may be subject to criticism. But I hope that even the critics of specific items will become convinced that this field of research does not belong to *explanation* but to a *sub-section of statistical inference*. It may be that this part of statistical inference has been neglected in the past since an adequate treatment requires the use of epistemic and pragmatic notions.

I shall first mention two ideas I accept from Gärdenfors. The first contains a liberalization and the second an improvement of corresponding concepts in Hempel's account of statistical explanation. Hempel, by parallelling the statistical with the deterministic case, required that explanations had to be potential predictions. From this, it immediately follows that the explanans must make the explanandum highly probable. Gärdenfors, on the other hand, accepts what could be called a "probabilistic minimum requirement" for statistical explanations. The basic idea is as follows: "the main purpose of an explanation is to give information about the explanandum in such way that it appears less surprising than before the explanation was given."[1]

This requirement for a *decrease* of the *surprise value* of the explanandum is tantamount to asking for a (non-trivial) *increase* of the explanandum's *belief value.*

I myself voted for an intermediate position in (10) when I argued that the explanatory argument has to satisfy what I called the "Leibniz-condition." This would amount to the condition that the statistical probability be greater than 1/2.

The improvement consists in replacing the notion of accepted sentences at a certain time by the concept of a knowledge situation. This replacement will have two main consequences. First, we shall have to do not only with *one* objective probability as in Hempel's account, but with a *whole family* of statistical probabilities. Secondly, by contrast to Hempel's procedure, we shall not keep the two kinds of probabilities separate, but rather we shall have to *combine* them in the literal sense that we will form out of them *probability mixtures.*

In (1)–(5) I shall try to formalize Gärdenfors's ideas, combining this formalization with slight modifications and improvements.

(1) *Knowledge situations: static description*

I begin with a possible world description of the knowledge situation of a person X at a time t. (On the whole I shall omit the reference to X and t

but I shall, like Gärdenfors, always presuppose that a knowledge situation is the total knowledge which a particular person has at a certain moment.) For simplicity, the *language* expressing the knowledge is chosen to be very primitive: the language contains one-place-predicates only and logical connectives but no quantifiers. Atomic sentences are either of the form Fa or of the form $p(G,F) \geqq r$, the latter meaning that the statistical probability of G relative to F be at least r.

The knowledge situation will be constructed by a *possible world model.* The basic idea is very simple. The knowledge of a person is fixed by the complementary concept, i.e., by what this person does not know. And what he or she does not know is determined by those possible worlds compatible with his or her knowledge. (For Hegel's World Spirit, *e.g.*, who, as I guess, has a total knowledge, the class of possible worlds compatible with its knowledge shrinks down to the unit class consisting only of the real world.)

Formally, a knowledge situation K is introduced as a quintuple:

D1 $$K = \langle U, W, B, \{I_w\}_{w \in W}, \{P_w\}_{w \in W} \rangle.$$

U, the universe of discourse, is a non-empty set of objects , which, for simplicity, we take to be the same in all possible worlds (or if you prefer Carnap's way of speaking, in all possible state descriptions). W, the set of possible worlds (of possible state descriptions) in question, will be called the *space of ignorance*. This name was chosen in order to remind us of the epistemic relativity of this set. Our person knows exactly those facts which are described by sentences that become true in all elements of W. Normally, a different knowledge situation will be characterized, among other things, by a different space of ignorance.

In addition to the first epistemic component W in K, we have as a *second epistemic component* a subjective (personal) probability function B, called the *belief-function.* It is a probability measure on the power set $\mathcal{P}(W)$ of W. (It should be mentioned in passing that, normally, B will not be a Laplacian probability, i.e. the possible worlds compatible with a knowledge situation are *not* assigned the same probabilities. If I assume that, presently, my friend Max is in France rather than in Germany, this means, technically speaking, that according to the knowledge I have, my belief-probability in the real state of the world, in which my friend is in France, is greater than the belief-probability that the real state of the world is among those in which he is in Germany.) Generally speaking, $B(V)$ expresses, for every subset V of W, the degree to which our person believes the real world being an element of V.

I_w is, for every $w \in W$, a semantic interpretation function. And P_w is, for every $w \in W$, a *statistical probability measure* on $\mathcal{P}(U)$. For the sake of simplicity, you may identify, for our primitive model worlds, each P_w with a relative frequency.

A brief technical remark on the use of the I_w's and P_w's will be in order. We construct these functions in such a way that the following holds:

(1) For every atomic sentence Fa,

$$I_w(Fa)=1 \text{ iff } I_w(\mathrm{a}) \in I_w(F);$$

otherwise it is 0.

(2) For every sentence $p(G,F) \geqq r$,

$$I_w(p(G,F) \geqq r)=1 \text{ iff } P_w(I_w(G),I_w(F)) \geqq r$$

otherwise it is 0.

(3) For a complex sentence s, $I_w(\mathrm{s})=1$ or 0 in accordance with the rules of propositional calculus.

By the *belief value of the sentence Fa* in the knowledge situation K we understand $B(\{w \mid I_w(Fa)=1\})$, *i.e.*, the degree of belief that Fa holds in the real world.

The phrase "*s is known in the knowledge situation K*" is short for the sentence "$\bigwedge w(w \in W \rightarrow I_w(s)=1)$" of our metalanguage, whereby W is the second member of K.

Exact knowledge of the relative frequency of Q can be expressed by means of the present formalism. It just means that for all $u, v \in W$, $P_u(I_u(Q)) = P_v(I_v(Q))$. (Actually, this statement not only says that the relative frequency of the individuals belonging to Q is the given one, but, in addition, that this is the relative frequency in all possible worlds compatible with what I know.)

(2) *Knowledge situations: dynamic description*

Suppose our person has a given knowledge, represented by K, which is increased by additional knowledge. The content of the latter is expressed by a set S of sentences. The class of possible worlds is then reduced to a class W_S and the original belief-function B is transformed into a function B_S. More exactly:

D2 K_S is the *enrichment* of the knowledge situation K by S iff

(1) $K = \langle U, W, B, \{I_w\}_{w \in W}, \{P_w\}_{w \in W} \rangle$;

(2) S is a set of sentences;

(3) $W_S = \bigcap_{s \in S} \{w \mid w \in W \wedge I_w(s) = 1\}$;

(4) $B(W_S) \neq 0$;

(5) B_S is that probability measure on $\mathscr{P}(W_S)$ which satisfies the condition: for all $V \subseteqq W_S$, $B_S(V) = B(V, W_S)$;

(6) $K_S = \langle U, W_S, B_S, \{I_w\}_{w \in W_S}, \{P_w\}_{w \in W_S} \rangle$.

In this definition, (3) expresses the requirement that in the new class of possible worlds all sentences of S must come out true. (5) says, first, that the new belief function is only defined on the restricted set $\mathscr{P}(W_S)$, and, secondly, that for any subset V of W_S, it gives the same result as the original function B, *given that S holds.* (4) formulates only the presupposition for building the conditional probability in (5)

(3) *Expected probabilities.* (*Subjective expectations of statistical probabilities constructed as "probability-mixtures".*)

In most cases a person does not know the exact value of the statistical probability (relative frequency). He or she will then have to *guess* this value. In such a situation we may 'weigh' our objective probabilities or relative frequencies by our subjective belief-probabilities. A bit more exactly: in order to determine in a given knowledge situation the probability that an object has property A, we construct a probability of second order, *i.e.*, we first calculate for each possible world w, belonging to this knowledge situation, the value of $P_w(A)$ and we then multiply this value with $B(\{w\})$, *i.e.*, with our subjective belief that w is the real world. This procedure is admissible since such a "probability mixture" is again a probability.

In this way we get a new probability measure P_w. Two generalizations suggest themselves immediately: (1) we may form B_V for every $V \subseteqq W$ with $B(V) \neq 0$; (2) instead of introducing the absolute probability $P_V(A)$ we immediately define the conditional probability $P_V(G,F)$.

Thus, let us be given the knowledge situation $K = \langle U, W, B, \{I_w\}_{w \in W}, \{P_w\}_{w \in W} \rangle$. For all $V \subseteqq W$ with $B(V) \neq 0$ holding, and for all predicates F and G, whose extensions are included in U, we define:

D3 $$P_V(G,F) := \frac{1}{B(V)} \sum_{w \in V} P_w(I_w(G), I_w(F)) \times B(\{w\})$$

For infinite V the sum has to be replaced by the integral

$$\frac{1}{B(V)} \int_V P_w(I_w(G), I_w(F))\, \mathrm{d}B(w).$$

Intuitively speaking, $P_V(G,F)$ is the *probability that an individual of F will be found to be in G, given that the real world* (or *the real state of the world*) *lies in V.*

In what follows we shall have to make use of P_V only for the special case $V = W$. It should always be remembered that we must distinguish between *three categories of probabilities*: the belief probabilities B; the objective probabilities P_w; and the probability-mixtures P_V with the limiting case $P = P_W$. We call the probabilities of the third class *expected probability measures* or *expected probabilities.*

Before continuing I shall briefly mention two respects in which the reconstruction of knowledge situations, following the suggestions of Gärdenfors, is superior to Hempel's concept of the class A_t of sentences accepted by a person X at a time t:

(i) While it is required that A_t is consistent and closed with respect to logical consequence, the question remains unanswered how probable or how improbable the person X considers sentences not belonging to A_t. In other words, the degree of certainty which is assigned to such sentences by X is not taken care of.

(ii) The enrichment of a knowledge situation in the sense of D2 may consist in the acceptance of statistical information, contradicting the person's calculated expected probabilities.[2] In such a situation we must remember the epistemic relativity of W. In the present case, W must be replaced by a smaller class W^* from which all those possible worlds are eliminated in which the new statistical proposition is false. (This will have the practical effect, that the second order probability distributions will be more concentrated on a definite value.)

(4) *Building Hempel's notion of maximal specificity into the concept of a knowledge situation*

I shall begin this section with my first critical remark: whatever result will be obtained at the end of our attempts, the explicated concept ought *not* to be called "*explanation.*" I have, in my introductory remarks, distinguished between three types of statistical single case substantiations of decreasing strength, *viz.*, the "high probability type" (Hempel), the type satisfying the Leibniz-condition (my earlier attempt), and the type involving a decrease in the surprise value (Gärdenfors). If it can be shown that not even type (1) may be called "explanation" then *a fortiori* the same holds for the other two types as well. My objection is a very simple one. Every attempt to interpret a probabilistic single case substantiation as an *explanation* founders on the possible realizability of the improbable event. Hempel, *e.g.*, and all those who followed his way of thought or tried to improve it, worked on the tacit premise that the event, whose occurrence was predicted with high probability, actually takes place. But it need not. Here, we come into an obvious conflict with an elementary adequacy requirement of explanations of events, namely that the event actually takes place and is not just a possible state of affairs. But if it does not take place we would have to say: "the following argument C explains why E. But unfortunately, E did not take place." This, of course, is absurd.

The deterministic case excludes such a situation. Here, only two "negative" cases may obtain. Firstly, the proposed cause A_1 may turn out not to be the real cause A_2 because A_1 has been "screened off" by A_2. this only shows that the suggested explanation was partially wrong whereas the explanandum event remained unchanged. Secondly, the deterministic prediction may fail: it is not the predicted event E, but something different which happens. *Prima facie*, this looks similar to our present case. But, still, there is a fundamental difference. In the deterministic case we would have a typical case of *falsification.* Some of the premises used in the argument must be wrong. No such conclusion is allowed in the present case: all the premises as well as the statistical argument may be correct, although E does not happen. Therefore, in contrast to the falsification case, we would not at all, in retrospect, say: "Our prediction was incorrect or irrational". Rather, we would say: "It *was* rational to predict and to expect the occurrence of E. But, unfortunately, the improbable event non-E took place."

These are my reasons for claiming that the "Hempel-Gärdenfors-line" *cannot* lead to a concept of statistical explanation. This does not imply, of course, that the explications suggested by their ideas are worthless. Quite the contrary is true. The concepts *are* of importance, although they do not belong to the domain of statistical *explanation* but to the domain of statistical *inference.* Their importance derives partly from the fact that they deal with a widely neglected aspect of statistical inference. Usually, this term is applied only when statistical hypotheses become the *objects* of critical studies, and not, like in the present case, when accepted hypotheses of such a kind are *applied* for predictive purposes. There is small wonder that this aspect is neglected outside philosophy. For its explication presupposes, as we have realized, epistemic concepts, like knowledge-situation, which are not to be found in technical literature on mathematical statistics.

Even if one accepts this argument, one may still raise the question whether it would be appropriate to apply the term "explanation" *in certain favourable cases.* My general attitude to this is that at this point we stop doing philosophy of science and enter a discussion which is partly psychological and partly linguistic in kind.

Let us now continue the formal explication. Our task is reduced to the problem of describing the pragmatic circumstances in which a statistical single case inference is permitted. As is well known, the main problem arising in this context is what Hempel called the "problem of maximal specificity". Following Reichenbach, he presented a solution according to which "a narrowest reference class" is to be chosen. We shall use this basic idea, but we shall, in order to remain in accordance with Gärdenfors's pragmatic notion of knowledge situation, replace the statistical probability by the *expected probability measure* P_W (in the sense of D3).

We split up our task into two steps. In the first definition we explain what it means for a knowledge situation to be a knowledge situation of maximal specifity. If the singular prediction is Fa, then it contains as an essential ingredient the claim that the belief value of Fa equals the expected conditional probability of F, given G, whereby G is the strongest projectible predicate (or: G is the smallest nomological class) of which one knows in this knowledge situation that it applies to a (or: that a is a member of it).

More exactly:

D4 K is *a knowledge situation of maximal specificity* iff

(1)$K = \langle U, W, B, \{I_w\}_{w \in W}, \{P_w\}_{w \in W} \rangle$;

(2) for every predicate F and for every individual constant a: if G is the strongest projectible predicate (or the narrowest nomological class) for which, in all $w \in W, I_w(Ga) = 1$ holds,[3] then $B(w \mid I_w(Fa) = 1) = P_W(F, G)$ (*i.e.* the belief value of Fa in K equals the conditional expected probability P_W of F on G).

It should not be overlooked that, via W, the "strongest predicate" G is *epistemically relativized*. After all, we do not consider *all* possible worlds, in which Ga holds, but only that subclass of them which is *compatible with* the knowledge situation K.

(5) *Statistical single case substantiation*

I should like to emphasize three peculiarities of the following definition.

First, seeming paradoxes will be avoided by distinguishing between *various knowledge situations* in *one and the same* objective situation. Secondly, the dynamic analyses of (2) will thereby help us to describe in precise terms how a given knowledge situation is transformed into a new one by adding additional knowledge. Thirdly, Hempel's requirement that the probability of the explanandum must be a high one will be liberalized to the more general requirement that the surprise value of the explanandum is lowered, *i.e.* that the additional information supplied by the explanans increases the belief value of the explanandum.

A formulation giving necessary conditions only I shall call a quasi-definition. We get the following quasi-definition:

D5 *X is in the knowledge situation K a statistical single case substantiation for E* only if

(1) $X = \langle K, E, T, C, K_{T \cup C}, K_E \rangle$;

(2) K is a knowledge situation of maximal specificity (in the sense of D4);

(3) E, the "explanandum," is a singular sentence (but its predicate may be complex in the sense of being formed with the help of connectives);

(4) T is a finite set of statistical sentences, the "statistical component of the explanans";

(5) C is a finite set of singular sentences, the "singular component of the explanans";

(6) K_E is the enrichment of K by E;

(7) $K_{T \cup C}$ is the enrichment of K by $T \cup C$;

(8) $T \wedge C$ is not known in the situation K_E (*i.e.*: $\neg \bigwedge w(w \in W_E \rightarrow I_w(T \wedge C) = 1)$;

(9) $B_{T \cup C}(\{w \mid I_w(E) = 1\}) > B(\{w \mid I_w(E) = 1\})$ (*i.e.*, the enrichment of K by the statistical information $T \cup C$ increases the belief value of E).

The three important items are the following ones: (i) the fact that K is of maximal specificity; (ii) that there exists an enrichment $K_{T \cup C}$; (iii) that the increase of the belief value is as described in (9).

As pointed out by Gärdenfors, the difference between the two belief values occurring in (9) may be identified with the explanatory power, or, as I would prefer to say, with the *substantiation power* (or: *argumentative power*) of $T \cup C$ relative to E.

The third knowledge situation K_E is needed only to formulate (8). And (8), again, is just a "trivialization obstacle," preventing non-informative explanations, like "E because E" (the set T may be empty).

One of the advantages of the present reconstruction is that examples which resisted adequate analysis within earlier frameworks can now be easily analyzed. Thus, *e.g.*, a satisfactory account can be given of the well known paresis example of M. Scriven.[4]

III. STATISTICAL ANALYSES

In this section I shall confine myself to some brief remarks on Salmon's account. I have three excuses for doing this: (i) Salmon has many interesting things to say on the connection between statistical relevance and causal relevance as well as on how to explicate the "screening-off" relation as a means to define causality. As the problems of causality lie beyond the scope of the present paper, I shall not deal with any of these questions. And for this reason alone I cannot, of course, do him real justice. (ii) An important concept used by Salmon is that of a *homogeneous* reference class. Like the concept of statistical probability itself, the analysis of this important notion belongs to foundational research in probability theory. Therefore, I shall just take this concept as available and shall not enter into a discussion of whether Salmon's explication of this concept is satisfactory or not. Intuitively, a homogeneous reference class is a class which cannot relevantly be subdivided. (iii) Like Hempel, Salmon needs the notion of accepted sentences at a given time. (He presupposes, *e.g.*, that a statistical law is available as an initial information.) For the same reasons as before, I would prefer, in all such places, to use Gärdenfors's possible world model of knowledge situations. In particular, the increase in knowledge which plays a great part within Salmon's account ought to be treated as an enrichment K_S of the original knowledge situation K (*viz.*, D2). Of course, the matter is greatly simplified if we follow Salmon and treat the statistical regularities involved as definitely known or accepted.

Unlike Salmon, I shall not use the phrase "statistical explanation." I shall rather speak of a *statistical analysis*, consisting of two components, an *analysandum* and an *analysans*. The analysandum contains the initial information. Its first component is a singular sentence $a \in F \cap G$. (This comes from a slight correction of Hempel's reconstruction of the why-question. We do not ask: "why is this *thing* a G?," but rather: "why is this F in addition a G?") Secondly, the statistical information, relating F and G by the law $p(G,F)=r$, is supposed to be given as well. (I accept this liberal assumption. No change, in principle, arises if we admit initial information not containing such a statistical law.)

Thus, the *analysandum* $\mathfrak{A}$ can formally be represented by:

$$\langle a \in F \cap G,\ p(G,F)=q \rangle$$

According to Salmon's basic idea, this initial information is improved in the following ways: (1) first, a minimal partition of F in, say n, homogeneous subclasses is performed:[5]

$$F \cap C_1,\ F \cap C_2,\ \ldots,\ F \cap C_n.$$

(2) Second, the probabilities of G relative to these classes as reference classes are determined. Supposing that we succeed in doing this we get n statistical laws: $p(G, F \cap C_k) = r_k$ $k = 1, \ldots, n$. (3) Third, we must find out into which of the subclasses the individual a belongs. If this is the class C_i, then the singular initial information is to be strengthened into: $a \in F \cap C_i \cap G$.

The sentences under (2) and (3) together make up what I call the analysans $\mathscr{L}$ and the transition from $\mathfrak{A}$ to $\langle \mathfrak{A}, \mathscr{L} \rangle$ I suggest calling a *statistical depth analysis* of the analysandum $\mathfrak{A}$.[6]

I must now explain why, again, it seems to me inappropriate to apply the term "statistical explanation" to this. The objection against the Hempel-Gärdenfors explication certainly does *not* apply in the present case, since Salmon presupposes that the explanandum *is known to be true.* However, the adequacy condition (2) is violated. In order to see this, we must remember that Salmon, unlike Carnap, takes as the explicatum for "degree of confirmation" not Carnap's c-function but Carnap's relevance measure. And as the particular relevance measure is always the result of a comparison of two probabilities, statistical explanations, in Salmon's reconstruction, are no arguments at all. However, it is not this non-argumentative interpretation of the notion of statistical explanation against which I will argue. Rather, my objection will be based on the closer inspection of the different kinds of relevance. If, in all cases, *positive relevance* would come out, everything would be alright. But, suppose that what I called the depth analysis results in negative relevance, we would get the following three results: $a \in F \cap C_i \cap G$; $p(G,F) = q$; $p(G, F \cap C_i) = r_i$ with $r_i \ll q$; and $q < 1/2$.

Checking the adequacy conditions (2) and (3) would lead to the following little game of question and answer:

(1) Question: Why is a which is an F in addition a G although the probability q of an F being G is not very high?

(2) Answer: Because we found out that a is, in addition, an element of C_i and this fact *lowers considerably* the probability of being a G, namely from q to r_i. Or, to reword it in Gärdenfors's terms: something is called an explanation in spite of the fact that its surprise value did not decrease but is considerably increased.

We cannot seriously call something an explanation if after this "explanation" has been given, we are much more surprised than we were before.

It is easy to see that in the majority of interesting cases it is again the realizability of the improbable which is responsible for this queer result.

Thus, we come to the conclusion that there is no concept of statistical explanation which satisfies the three requirements of adequacy. The presystematic notion of statistical explanation must be split up into two different concepts. The first one is the *concept of single case substantiation.* It is normally applied for predictive purposes and, therefore, does not presuppose that the predicted event really will occur. (If it occurs, then the surprise value is decreased by the substantiation.) The analysis of this kind of argument should be considered as a subsection of statistical inference. The philosophically important difference between this and that part of statistical inference which deals with the support and test of a statistical hypothesis, lies in the fact that a satisfactory account of the application of the statistical hypotheses in question requires a careful epistemic analysis of various knowledge situations involved. The second concept is that of a *statistical depth analysis.* It can be performed only after the singular event in question has occurred. What we gain by this analysis is an *understanding* of the working of the underlying statistical mechanism by which the event was brought about. This gain of understanding need not be combined, as in the previous case, with an increase in the belief value of the event. It may very well happen that, after we have received sufficient information about the statistical mechanism and, in addition, after we have learned to which homogeneous subclass our event belongs, we will be more surprised than we were before.

IV. INTUITIVE VERSUS FORMAL ACCOUNTS

There is still an open question, namely the dilemma (*A*): How can our result be reconciled with the fact that, on an intuitive level, we can give countless examples of statistical explanations?

I shall say only a few words about this problem. For whatever answer one may give, contributing to a clarification of this situation, it will no longer belong to philosophy of science. One may regard it as belonging to the psychology of the presystematic use of metascientific terms.

First, we must never forget that the word "explanation" has many different uses. It can be applied in both types of cases, although in the first type of case it gives us a very different kind of information than the single case substantiation. What is explained, in the Hempel-Gärdenfors account, is not why the event *E* will happen but *why it is rational to expect* the occurrence of *E*. If, then, *E* really happens, everything seems to be alright. In such cases the phrase "*E* is explained" may be taken just as an abbreviation for what, in a more subtle analysis, ought to be rendered as "*E* was rationally to be expected and, besides, *E* actually occurred." There is a decisive difference between this and the deterministic case where we have a deductive argument. Here, the occurrence of non-*E* amounts to a proof of the falsity of at least one premise. This is not so in the present case. Suppose the probabilistic reasoning to be correct, we should say: "non-*E* happened although it had been rational to expect the event *E* to happen." That a rational expectation is disappointed does not imply *that the rationality of the expectation is obliterated.*

Secondly, let us suppose that from Hempel's point of view as well as from Salmon's point of view a situation is realized which is most favourable for applying the term "explanation":

(a) Before the time of the event a maximal specificity argument with high probability for E is provided[7];

(b) at t E really happens;

(c) a retrospective analysis of the Salmon-type gives high positive relevance.

Prima facie there seems to be no reason why not to apply the term "explanation" in such a case. The reason why I would hesitate to follow such a suggestion is the following one. It makes the correctness of the application of the term "statistical explanation" dependent on chance. On the other hand, a term like "explanation" is used by us as an achievement term. I shall illustrate the situation with a simple analogue from ethics. Suppose two persons have the same good intentions but only the first one is able to realize them while the second one, by bad luck, misses the realization. We would then not say that the first one was "better" than the second but only that he was the luckier one of the two.

We may apply this to the present situation. Suppose that two persons X and Y in a similar situation have performed all the intellectual achievements as described in (a) and (c). But only in the one case E happens, while in the other case it does not. If the terminological suggestion mentioned is accepted we would have to say: *X was able to give* an explanation whereas *Y was not able to give* one. This, again, sounds as if X would be *better* than Y, though, of course, not morally, but with respect to his intellectual achievement. *But this is just not true. X was not any "better" at all, he was only luckier than Y.*

This is the reason why I come to the conclusion that within systematic contexts the notion of statistical explanation ought to be given up in favour of the two suggested successor concepts, the concept of *statistical single case substantiation* on the one hand and the concept of *statistical analysis* on the other.

NOTES

1. Gärdenfors [1], p.1.
2. In Hempel's case, of course, we should have to observe a conflict between a newly accepted probability statement and accepted old ones.
3. This phrase only expresses the statement that Ga is known to be true in the knowledge situation K. Therefore, G is the *smallest* nomological class for which this holds.
4. For a detailed discussion, see Gärdenfors [1], IV.
5. The requirement of minimality is needed in order to prevent "vacuous" subdivisions of homogeneous classes into subclasses for which the conditional probability of G will be the same.

6. In Stegmüller [12] I added "of minimal form" because the result of the analysis can in principle be improved, in particular in such cases where F is an attribute of a whole "family of attributes" in the sense of Carnap. But since all such further improvements are irrelevant for our knowledge about a I now drop this qualifying attribute.

7. In the present context it does not matter whether we refer back to Hempel's or to Gärdenfors's reconstruction.

REFERENCES

1. Gärdenfors, P., *A Pragmatic Theory of Explanation*, Lund, Sweden, Working Paper No. 19, 1976.
2. Hempel, C.G., *Aspects of Scientific Explanation*, New York/London, 1965.
3. Hempel, C.G., "Maximal Specificity and Lawlikeness in Probabilistic Explanation," in: *Philos. of Sci.* 35 (1968), pp. 116–133.
4. Jeffrey, R.C., "Statistical Explanation versus Statistical Inference," in: Salmon, W.C. (ed.), *Statistical Explanation and Statistical Relevance*, pp. 19–28.
5. Niiniluoto, I., "Inductive Explanation, Propensity and Action," in: Manninen, J. and R. Tuomela (eds.), *Essays on Explanation and Understanding*, pp. 335–368.
6. Salmon, W.C. (ed.), *Statistical Explanation and Statistical Relevance*, Pittsburgh 1971.
7. Salmon, W.C., "Statistical Explanation," in: Salmon, W.C. (ed.), *Statistical Explanation and Statistical Relevance*, pp. 29–87, "Postscript 1971", pp. 105–110.
8. Salmon, W.C., "A Third Dogma of Empiricism," in: Butts, R.E. and J. Hintikka (eds.), *Basic Problems in Methodology and Linguistics*, Part Three of the Proceedings of the Fifth International Congress of Logic, Methodology and Philosophy of Science, Dordrecht 1977, pp. 149–166.
9. Scriven, M., "Explanation and Prediction in Evolutionary Theory," *Science* CXXX, (1959) pp. 477–482.
10. Stegmüller, W., *Wissenschaftliche Erklärung und Begründung*, Berlin-Heidelberg-New York, 2nd ed. 1974.
11. Stegmüller, W., *Personelle und statistische Wahrscheinlichkeit*, first half-volume: *Personelle Wahrscheinlichkeit und rationale Entscheidung*, Berlin-Heidelberg-New York 1973.
12. Stegmüller, W., *Personelle und Statistische Wahrscheinlichkeit*, second half-volume: *Statistisches Schließen – Statistische Begründung – Statistische Analyse*, Berlin-Heidelberg-New York 1973.
13. Stegmüller, W., "The Problem of Induction: Hume's Challenge and the Contemporary Answers," in: Stegmüller, W., *Collected Papers on Epistemology, Philosophy of Science and History of Philosophy*, Vol. II, Synthese Library 1977, pp. 68–136.
14. Stegmüller, W., "Carnap's Normative Theory of Inductive Probability," in: Stegmüller, W., *Collected Papers on Epistemology, Philosophy of Science and History of Philosophy*, Vol. II, Synthese Library 1977, pp. 137–149.
15. Stegmüller, W., *The Structuralist View of Theories. A Possible Analogue of the Bourbaki-Programme in Physical Science*, New York 1979.
16. Suppes, P., "Probabilistic Inference and the Concept of Total Evidence," in: Hintikka, J. and P. Suppes (eds.), *Aspects of Inductive Logic*, Amsterdam 1966, pp. 49–65.

17. Suppes, P., "Popper's Analysis of Probability in Quantum Mechanics," in: Schilpp, P.A. (ed.), *The Philosophy of Karl Popper*, Vol. I, La Salle, Ill., 1974, pp. 760–774.
18. Suppes, P., "New Foundations of Objective Probability: Axioms for Propensities," in: Suppes, P., Henkin, L., Moisil, Gr.C. and A. Joja (eds.), *Logic, Methodology and Philosophy of Science*, Vol. IV, Amsterdam 1973, pp. 515–529.
19. Wright, G.H.v., *Explanation and Understanding*, London 1971, pp. 13–15.

PATRICK SUPPES

(*Stanford University*)

SOME REMARKS ON STATISTICAL EXPLANATIONS

There are many persons so far apart philosophically that it is difficult for them to establish a common frame of reference and a common vocabulary as a basis for discussion. Other individuals can differ but operate within a common framework of concepts and a generally accepted approach to philosophical topics. Professor Stegmüller and I are of the latter kind. I have several criticisms of his views on statistical explanation and I want to make those criticisms in as sharp and definite a form as I can, but I also want to emphasize how much we agree in general philosophical approach. My many points of disagreement are evidence, I would take it, that the foundations of statistics and the foundations of the theory of belief involving partial or incomplete knowledge are still very far from a satisfactory formulation.

I have organized my remarks under six headings.

1. STATISTICAL EXPLANATION

I am not happy with Stegmüller's reticence regarding explanation. Such explanations are common in science, especially in the science of the last few decades, and it gives philosophy an idiosyncratic character in relation to the main trends of scientific discussion to deny the standard character of statistical explanations. I hold that there are three kinds of statistical explanation, and this threefold character, of which I have spoken elsewhere in the past, is important for the discussion of the problems we are dealing with today.

One kind of statistical explanation is theoretical in character. A good example would be the explanation of the analysis of radioactive decay, or, more generally, the kinds of explanations of phenomena given by quantum mechanics. Quantum mechanics is surely the most important empirical theory of the 20th century and is replete with what I would call statistical explanation of natural phenomena at the theoretical level.

The second kind of statistical explanation is experimental explanation.

G.H. von Wright (ed.), Logic and Philosophy, 53–58.

Here I would cite as a good example the extensive medical experimentation in which the theory of the phenomena studied is very unsatisfactory or essentially nonexistent — in particular, empirical studies of drug interactions, almost all of which have little theoretical basis.

The third sort of statistical explanation is more a focus of Stegmüller's discussion — explanation of our beliefs, for example, my belief as to why it will rain tomorrow.

2. EVENTS AND KINDS OF EVENTS

One distinction that is missing from Stegmüller's discussion is absolutely fundamental to the application of probabilistic notions in science. It is that between things and kinds of things or between events and kinds of events. Stegmüller's talk about necessity, as in his first requirement that the event must occur, is a demand to deal with events. But when we are discussing probabilistic theories of phenomena, for example, the kind of explanation given in quantum mechanics at the theoretical level, we are concerned with kinds of events, not events. Moreover, in the discussion of experiments we are also concerned with kinds of events, not events. The developed theory of modern statistics for handling this distinction is the theory of sampling. In testing a probabilistic hypothesis or theory we are ordinarily not interested in the occurrence of a particular event on a particular occasion but with the assignment of probabilities to kinds of events. This point needs elaboration but cannot be gone into in further detail here. I do want to insist that when we look at theories that are probabilistic in character, their formulation is in terms of kinds of events.

3. HIGH PROBABILITY FOR PREDICTED EVENTS

I certainly agree with Stegmüller about the absurdity of Hempel's requirement that predicted events have high probability. I would like to make a general remark on this matter.

It has sometimes been said that there are two strains of medieval thought, the Aristotelian and the Archimedian, and that we can trace these two traditions more or less continuously from Hellenistic times. The same dichotomy seems to exist in the theory of probability and its application to the real world. We have a tradition of philosophical discussion that seems to me rather remote from what might be called the Archimedian tradition of detailed mathematical analysis of real phenomena. One of the things that I am most critical about in Hempel's discussion — and I am critical insofar as Stegmüller is depending on this view — is that no serious contact is made with the many sophisticated, interesting and powerful examples of applying probabilistic notions to the study of natural phenomena.

Even limited contact with the classical works of Laplace and others shows how inappropriate it is to require that predicted events have a high probability. The reason is evident. The powerful applications of probabilistic notions in the real world have mainly used continuous probability distributions in which the events in question have probability zero at the most detailed, idealized level of observation. Here I agree with Stegmüller, but I am unsatisfied and unhappy with his introduction in this context of the notion of surprise. We had some discussion of Shackle's notion of surprise in Hilpinen's paper two days ago. I stated then in the general discussion my own view that the theory of surprise is a bad theory. I have not changed my mind since Tuesday. (For a detailed written opinion, see Luce and Suppes, 1965.) I do not want the application of probabilistic notions to the real world in any sense to rest on the very undeveloped and very unsatisfactory notion of surprise. When we talk about the successes of quantum mechanics we can look from one page of one treatise or one article to the next without finding any discussion of a notion or surprise. Of course, we can agree in cocktail-party talk that it is surprising that quantum mechanics is so unbelievably successful in analyzing natural phenomena. But I say it is not part of the serious application to the real world of a probabilistic theory such as quantum mechanics to have at hand a notion of surprise. What we have at hand and what we use extensively are the standard notions of testing a statistical theory. We have an elaborate technology for making such tests. It is important in the discussion of explanations to deal with that elaborate development and the way in which elaborate theories of a probabilistic character are tested or should be tested.

4. MAXIMUM SPECIFICITY AND TOTAL EVIDENCE

Let me begin this point by referring to an earlier fantasy given us by Laplace. This was his definition of probability. Laplace begins his treatise with the definition of probability in terms of the ratio of favorable cases to possible cases, but a few pages later, where serious examples are cited, the definition plays no important role.

A similar sort of fantasy has been engaged in by Reichenbach, by Hempel, and to some extent now by Stegmüller in the advocacy of maximum specificity and homogeneous reference classes. This advocacy is a serious mistake as an approach relevant to the application of probabilistic notions in developed science. To make a radical statement on this point, we can look in the last year's *Physical Review*, the largest physical journal in the world, and of the 3,000 to 4,000 published pages, something like 2,000 to 2,500 will be experimental articles using some statistical apparatus. My claim is this: Not one single article will make a reference of any kind to homogeneous reference classes or to maximum specificity.

There is a clear and definite Bayesian answer to the problem of maximum specificity and the homogeneous reference class, and I might say something about it. There is another title under which this problem is spoken of in philosophical circles, that of *total evidence*. I shall illustrate how from a Bayesian standpoint the problem does not arise in the first place. It is, so to speak, a nonproblem. The reason it is a nonproblem is that what corresponds to the problem of maximum specificity is simply for the Bayesian a part of the requirement of coherence. Suppose that I want to give you my probability that it will rain tomorrow (R) and I say, "Well, there are heavy clouds now" (H). If I believe $P(H)=1$, then

$$P(R)=P(R|H),$$

because by the elementary theorem on total probability

$$P(R)=P(R|H)P(H)+P(R|\text{not } H)P(\text{not } H),$$

and the assumption is that $P(\text{not } H)=0$. If you bring in another fact, for example, the barometer is dropping, then I simply introduce that and assign my probability to it in the standard way. In order to have a probabilistic coherent set of beliefs I must take account of information about the barometer. If I do not have that information, then my belief in the probability of rain may change; but if I already know about it, it does not change my probability. (For detailed treatment of these matters, see Suppes, 1966.) You can, if you will, bring in anything you like, but it will usually have no practical effect on my belief, because I have already absorbed most of the relevant information. So the Bayesian has a natural way of dealing with the problem of maximum specificity. It is a nonproblem, taken care of by the theorem of total probability and the requirement of coherence.

5. EXPLICIT KNOWLEDGE FRAMEWORK

I do not believe in the kind of explicit knowledge framework that Stegmüller has introduced us to and that he has regard for. I have my own addiction to explicit set-theoretical formulations, but when it comes to the subject under discussion here, I want to be very careful not to make matters explicit that should be left implicit. Induction, if you will, is like seduction on this point: It is often a mistake to be too explicit. A good example of this is to be found in standard statistical practice.

One difference of conceptual framework that has not been noted often enough in the use of probability concepts by statisticians, on the one hand, and by philosophers, on the other, is that philosophers are always trying to make everything explicit. They want to talk, as Stegmüller has, of an explicit knowledge framework. Another way that is a little more standard in probability theory is to talk in terms of a sample space or a probability space; but in the real world — and by the real world here I mean the real world of science — that is not the height of fashion. Sophisticated talk about applications of probability theory is like sophisticated seduction. It is always implicit what is beneath the surface. Sophisticated probabilistic talk in science is talk about random variables. Random variables are, from a technical standpoint, measurable functions defined on a sample space. So it seems easy to say: "Aha, there is a sample space there after all and it is just that the statistician has slighted it." But that is a mistaken way of looking at how things are done, because construction of an actual sample space is of no interest whatsoever. There is in statistics a theorem, sometimes called the theorem for unconscious statisticians, that imposes consistency conditions on a family of random variables in order for there to be a common sample space on which they can be defined. But no one in applications is ever concerned about this theorem. No one, above all, is in the least interested in exhibiting a suitable sample space. What one does is deal only with random variables. Moreover, the important and common concepts of mean, variance, and covariance apply only to random variables and their distributions, not to events of a sample space. (For amplification of these remarks about random variables, see Suppes, 1974.)

There is a deeper Bayesian point as well about these matters. It is that we do not want to be committed to some particular knowledge framework but want to express our viewpoint at a given point in time by a prior distribution over all our beliefs. Fixing an explicit knowledge framework in the case of beliefs is unrealistic and therefore mistaken. An explicit framework is justified in the case of theory, and sometimes in the case of experimentation. But even in the case of the statistical analysis of relatively simple experiments, the kind of first-order logic apparatus that Stegmüller has introduced does not adequately deal with the simplest kinds of applications that require relating tests of a statistical hypothesis to the theory of sampling. At this late stage in the development of modern statistics, it is rather like saying, "Let's analyze physics, and especially 20th-century physics, in first-order logic."

6. INTENTION AND ACHIEVEMENT

I take up briefly Stegmüller's remarks about achievement. First, when we are dealing with an individual case of prediction, what is fundamental is the result and not the intention. If your prediction is bad, that is tough luck. A theory of survival requires that we think in terms of results and not intentions. Suppose we own a firm whose market is highly sensitive to seasonal weather conditions. We think we run the firm well, but we encounter seasonal problems and the firm goes bankrupt. The fundamental fact is that we failed. Maybe our intentions were laudable, but we got into trouble nonetheless. I think that this is the way we have to think about the real world.

In a more general way, and related to what I said earlier about the concept of surprise, I do not think talk about achievement is of real importance within a scientific framework. The distinction we may want to make between intention and achievement is not an important problem in the case of sophisticated scientific applications of probabilistic notions. For example, when we try to explain from fundamental quantum mechanical principles the exponential law for radioactive decay, we do not leave anything to luck. Whether or not an individual atom decays at an individual instant is a very random affair obeying an exponential probability law, but whether the totality of experimental data fits the exponential law to the finest determinable degree is not at all a matter of randomness but a matter of careful and detailed experimentation which in the end leaves nothing to chance or to failed intentions.

REFERENCES

Luce, R.D. and Suppes, P., "Preference, Utility and Subjective Probability," in R.D. Luce, R.R. Bush, and E.H. Galanter (eds.), *Handbook of Mathematical Psychology* (vol. 3), Wiley, New York 1965, pp. 249–410.

Suppes, P., "Probabilistic Inference and the Concept of Total Evidence," in J. Hintikka and P. Suppes (eds.), *Aspects of Inductive Logic*, North-Holland, Amsterdam 1966, pp. 49–65.

Suppes, P., "The Essential but Implicit Role of Modal Concepts in Science," in K.F. Schaffner and R.S. Cohen (eds.), *PSA 1972*, Reidel, Dordrecht 1974, pp. 305–314.

WOLFGANG STEGMÜLLER

(*University of Munich*)

COMMENT ON "SOME REMARKS ON STATISTICAL EXPLANATIONS" BY PROFESSOR SUPPES

It seemed to me necessary to comment on Professor Suppes's comment, mainly for two reasons. First, a good many of his criticisms are, in my opinion, based on misunderstandings (for which, to some extent, I may be responsible). Second, it seems to me that it will be very difficult for most readers on simply reading my article and Professor Suppes's comments to find out what exactly is the difference between our positions with respect to the topic under discussion. If I am right it is not so much a contrast of views on particular items, nontechnical or technical ones, as a difference in philosophical attitude to questions in the philosophy of science. I therefore hope that some additional remarks will help to clarify the issue.

I shall subdivide my comment into six parts. First, I shall point to those items which are based on misunderstanding. Second, I shall try to localize our difference in opinion. Third, I shall mention an aspect of Suppes's account which I do not understand. Fourth, a few additional words will be devoted to my negative philosophical result. Fifth, the positive result of my paper will be illustrated with an example. Sixth, I shall give a summary of the philosophical issue as I see it.

I

I ought to have stated more explicitly that I shall be dealing with explanations of *events* only. Thus, the quantum mechanical explanation of radioactive decay was not my concern at all. I know that there are philosophers who entertain the view that a satisfactory explication of explanation has to cover both cases, events and laws or kinds of events. In my opinion such a belief rests on an illusion. Even in the deterministic case it is not the normal situation that a so-called explanation of a law by a more comprehensive theory consists merely in a deduction of the former from the latter. Take, *e.g.*, the explanation of Kepler's laws by Newton's theory. It must be part of this explanation, if it is adequate, that Kepler's laws are *false*, when viewed from the standpoint of Newtonian theory;

G.H. von Wright (ed.), Logic and Philosophy, 59–69.

therefore, a deduction is out of question. The explanation has, presumably, to be reconstructed as an *approximative reduction* of Kepler's theory to Newton's (whereby it is the latter, *i.e.*, the *reducing* theory which has to perform the approximation, since no "planet particles" with zero mass can be models of Newton's theory).

Another misunderstanding has to do with my use of the terms "surprise value" and "achievement." Since the latter occurred only in the last few lines of my paper in which I appealed to myself as a competent speaker I shall postpone this point for the moment.[1]

The term "surprise value," which occurred first in a literal quotation from Gärdenfors, has no systematic import at all, neither in Gärdenfors's paper nor in my reconstruction of his ideas. I could have omitted it entirely. Its whole use consisted in no more than in an intuitive or presystematic justification for the liberalization of Hempel's narrow conception of explanation. This liberalization, by the way, is emphatically accepted by Professor Suppes.

Suppes rejects the concepts of maximal specificity and homogeneous reference classes. The term "maximal specificity" is Hempel's and not mine. And the term "homogeneous reference class" is Salmon's and not mine either. It was not my intention to justify their ideas. Quite to the contrary. The *philosophical* upshot of my paper was *totally destructive.* But I concede that I may have confused my audience by my twofold strategy. The destructive *philosophical* intention was paralleled by a constructive *technical* suggestion. This suggestion consisted in a replacement of the notion A_t (the class of accepted sentences at t) by K_t (knowledge situation at t). While the former is exposed to the objections (i) and (ii) mentioned on p. 43 of my paper, the latter is not. But it could, of course, have been objected that one does not see how this new concept may be used in formal reconstructions. In order to exemplify this and thereby meet such a possible objection, the concept of statistical explanation of an event was, *as an exercise*, explicated for a simple model language. Here I partly followed Hempel and partly Gärdenfors. Thereby my own position remained neutral with respect to *other* technical details of, *e.g.*, Hempel's definition. If one thinks that Hempel's account ought to be modified then I hope the exercise will suffice to show how to perform the details of this change by simultaneously replacing A_t by K_t.

With respect to homogeneity, the situation is even simpler. In order to formulate the main objection against W. Salmon's account of statistical explanations of events, I had to sketch the formal structure of his explication, where essential use is made of this notion. My objection is independent of whether one finds the use of this notion disagreeable, as Suppes does, or whether one finds it appropriate.

II

In order to locate the difference between Suppes and me the distinction between *general* and *special philosophy of science* is important. Since I have dealt with this dichotomy at some length elsewhere,[2] I shall here restrict myself to a few remarks. Practically all of the earlier empiricist philosophers believed in a *general* philosophy of science in the following sense. According to their common conviction, all metascientific key expressions, like "is a law," "is a disposition," "is confirming evidence for," "is a theory," "is a (deterministic) explanation" are to be defined in general terms, without reference to particular scientific theories or their expositions and without reference to particular historical periods. The advocates of a *special* philosophy of science, as I shall call them, do not believe in this kind of philosophical undertaking at all. According to their opinion, all such terms can be explicated only in the context of particular theories.

While most of the philosophers holding the latter view are more or less historically oriented, like Kuhn and Feyerabend, Suppes is one of the few philosophers whose systematic interests concentrate on questions belonging to special philosophy of science. Most of his works belong to this field. I myself started as a convinced proponent of general philosophy of science, but during the past years my position has become much closer to that of Suppes. For instance, I no longer believe that there is one single important metascientific notion which can be explicated satisfactorily in general terms.

However, there still seems to be an essential difference between our views in this respect. I am much less skeptical than Suppes about the possible accomplishments of a general philosophy of science. It may formulate a general framework for special research; or give some important necessary conditions (although not necessary *and sufficient* conditions) for a metascientific concept. It may even produce important contributions to the philosophy of language or to metaphysical issues, like the idealism-realism controversy, as is illustrated by the writings of W.V. Quine and H. Putnam.

I admit that I *may* be wrong in my optimism. But if I am not, then it is in principle admissible to make use of miniature models expressed either in a natural language or in simple formal languages. It is admissible if it helps us to new insights. In particular, it is admissible if the philosophical result we obtain is not a positive one but, as in the present case, is mainly destructive (see [V] below).

III

When Suppes published his Bayesian alternative to Carnap's well known methodological principle and to Hempel's requirement of maximal specificity, he was a convinced proponent of subjectivism in probability theory. He has evidently changed his mind since then. In his Bucharest paper [18], pp. 526–528, he showed how to introduce *statistical probability* as a *theoretical entity*, thereby making Popper's original concept of propensity more precise. It is known to me, of course, that not all Bayesians are subjectivists. On the other hand, there is no question that the subjectivist's position becomes much more difficult when he decides to become a probabilistic dualist without giving up his Bayesian point of view.[3] Among other things, he will be confronted with the question of how he gets the *a priori* knowledge about the world to start with. But it is not my intention here to make critical remarks belonging to the area of foundational research in probability theory. My present concern is *whether a Bayesian dualist can avoid working with probability mixtures.* If he cannot, I am satisfied. If he thinks he can, then I do not understand his position. It is a pity that Suppes did not comment on the question of probability mixtures, since this might have clarified this important point. As things stand, I actually do not know whether there really *is* an effective alternative to my suggestions, nor do I know what it would look like. Perhaps our positions are not at all as far apart from each other as Suppes's comments might suggest.

In the context of criticizing the requirement of total evidence Suppes quotes his paper from 1966 where he dealt with this point, and related matters, in detail. The quotation may create the impression that I was not familiar with Suppes's article. I mention in passing that this would be a mistake. In [10], pp. 684ff., I devoted two sections to this approach. In [12a] I tried to give a correct description of it, followed by a critical evaluation in [12b].

IV

I shall now add a few words to the philosophical, *i.e.*, to the *destructive* part of my paper. All attempts at explication known to me either interpret statistical explanations of events as specific arguments or as something which has no similarity to an argument at all, even if the word "argument" is taken in its most embracing and most liberal sense.

To the first type of attempts belongs Hempel's account as well as Gärdenfors's modified and liberalized sketch which I tried to formalize above. In my opinion, the following holds: whatever refinements, technical improvements and generalizations one adduces to an explication of the argumentative kind, all one can thereby achieve is a form of argument which may be used for *predictive* purposes only.[4] Take *e.g.*, the following quotation from Gärdenfors:

> We can ... distinguish between two knowledge situations in connection with an explanation of a sentence "*Qa*", In the first situation you know that "*Qa*" is true, but you do not know why. In the second you do not know that *Qa*, nor do you expect that *Qa*. This situation is normally the knowledge situation you were in before you discovered that *Qa*. The definition of explanation which will follow will be based on the second of the knowledge situations, the one where you do not know that *Qa*.[5]

And at a later place, having reformulated the problem in his terms, he says: "This problem is the classical problem of 'single case probabilities' translated into the framework of knowledge-situations ..."[6] Do not these remarks show that a restriction to *predictive* cases is intended? If not, what should we say if non-*Qa* happens although a correct "explanation" of *Qa* had been given? Are we to say: "I explained why *Qa* but, unfortunately, *Qa* did not happen"? I consider this an absurdity. One who thinks it is not has to show why it isn't.

It seems to me that the position taken by Gärdenfors (and, of course, by Hempel and many others) is due to the engrained intuition that explaining an event amounts to the same as showing why this event was rationally to be expected. But this intuition is wrong if the rational expectation rests on statistical laws.[7] If an explanandum *E* does not occur then the explanation of *E* must be wrong. On the other hand, the non-occurrence of *E* does not prove that the prediction of *E* was not rational.

There *is* a sense of "explication," different from Carnap's, in which we can speak of explicating explanation. I shall call it the *little story explication of a term.* In this sense "I explained *E*" (or: "I explained why *E*") is just an abbreviation for a longer story, namely the following one: "first, I explained why it was rational (or at least not irrational) to expect the occurrence of *E*; *and*, secondly, *E actually occurred*." This little story talk is the only way of *solving* the dilemma (*A*) I know of, if we start with the argumentative view of explanation. But it should not be forgotten that it does not depend *only* on whether we are able to tell the story. Chance might have made it impossible for us to tell *this* story, thereby preventing us from "giving an explanation" of the kind mentioned.

But we need not start with the argumentative view at all. If we are primarily interested in what Gärdenfors in the quoted passage called the "first situation," *i.e.*, if we take it for granted that the event to be accounted for is a *fact*, then the increase of information looked for will be adequately analysed in a Salmon-type explanation. In the terminology which I suggested, this change should be formulated in the following way: our original interest *in an applied statistical inference for predictive use* is now replaced by an interest of a very different kind, namely by the search for *a statistical mechanism underlying an event which has been realized* or, briefly, by the search for a *statistical analysis.*

In this case, an explication of "explanation" *in the little story sense* can be given. In order to parallel this case to the former, we decide to use the word "statistical analysis" in a narrower sense than in Part III above, namely by leaving out the additional information telling us to which particular subclass a belongs. And by "statistical explanation" we can understand the conjunction, consisting of this analysis (which, be it noted, contains among other things a subdivision of E into n subsets and n probabilistic laws with definite value r_k) *and* the singular statement telling us into which class a belongs, *i.e.*, $a \in F \cap C_i \cap G$.

Prima facie, it looks as if this reconstruction of "statistical explanation" is superior to the first one (in the little story sense). For, in the first case, as we have seen, it depends on chance whether we can tell the story at all (if E does not happen we cannot). This time, we can in principle always tell a story of the outlined structure. But another disadvantage now lies in wait for us. *It depends on mere chance whether one can claim that an explanation is an answer to a why-question*; in case of irrelevance or negative relevance one cannot (about this, see p. 48).

This is the place to comment on my use of "achievement." Actually, the use of this term was part of an answer to a question raised elsewhere by I. Niiniluoto.[8] He asked why the term "statistical explanation" could not be reserved for the Salmon-type *with positive relevance obtaining.* To illustrate, take the example given by Salmon in [7] on p. 208. There, a mixture of uranium 238 atoms (whose half-life is 4.5×10^9 years) and polonium 214 atoms (whose half-life is 1.6×10^{-4} years) is considered. Suppose that within some small specified time interval a decay occurs. Although there is a very high probability of a polonium atom disintegrating within that interval, a given disintegration *may be* of an uranium atom. Suppose both persons X and Y (our "explainers") give a precise account of the situation. But only X encounters a polonium atom disintegrating, while Y, by chance (or as one should say: by misfortune), encounters an uranium atom disintegrating. I said "by misfortune" because, if we follow Niiniluoto's terminological suggestion, we should have to say *that X was able to give an explanation while Y was not able to give one.* This sounds very inappropriate to my ears because the intellectual accomplishments of the two people are of exactly the same value and it is only chance which differentiates between the two cases.

Thus, from a logical point of view, the situation seems to be clear enough. There are important reasons to distinguish, conceptually and terminologically, between the *predictive use* of statistical inference and the use of statistical information in order *to get an understanding* of the mechanism underlying a fact. And there are equally strong reasons for calling none of them an explanation.

It is true that in the little story sense of explication one may continue to use the phrase "statistical explanation of an event." One can even choose between two alternatives. But one ought to be aware that one will have to pay a price in either case. If one is willing to pay in at least one of them, everything is right. If one pays in both of them one must be careful to avoid ambiguities. One has then gained two advantages. First, one has salved one's conscience which forced one to squeeze a meaning out of the phrase "statistical explanation of an event." And, second, one may cheerfully say to oneself that one knows how to overcome the dilemma (*A*).

V

I now return to the constructive aspect of my paper. Instead of describing the situation in general terms (which I could do only by rephrasing what I have already said), I shall illustrate it with a particular example, namely the *paresis example* of M. Scriven.[9] This example has attained notoriety because it has turned out to be very recalcitrant to adequate analysis.

In reconstructing it as a case either of rational prediction or of rational explanation (in the first "little story" sense of explanation), let us assume as general premises the following three statements (*a*)–(*c*) and (*d*) as an additional singular premise:

(*a*) Paresis develops only in patients who have been syphilitic for a long time.

(*b*) Only a small number of syphilitic patients will ever develop paresis.

(*c*) No other factor besides the one mentioned in (*a*) is known to be relevant for the development of paresis.

(*d*) Person *p* has been suffering from syphilis.

Suppose *p* actually developed paresis. Scriven maintains, contrary to what follows from Hempel's theory, that (*a*)–(*d*) explain why *p* developed paresis. Scriven himself appeals to the more general rule that if the property *R* is *the only known cause* of the property *Q*, then one can explain why a certain individual has the property *Q* by pointing out that he or she has the property *R*. Hempel rejects this by saying that "a condition that is nomically necessary for the occurrence of an event does not, in general, explain it."[10] And he defends his own theory by the following counterexample which, in his opinion, has the same structure as the paresis example. No one wins the first prize in the Irish sweepstake without buying a ticket. But only one of those who have bought a ticket wins the first prize. Hempel claims that we cannot explain why someone wins the first prize by pointing out that he has bought a ticket.

Many readers will have the feeling that Scriven is "somehow" right and that the counter-example is not quite correct, without being able to say why this is so. It is an easy matter to put things right, first, by using an appropriate knowledge situation to reconstruct the background knowledge in question, and second, by making plausible empirical assumptions.

In the knowledge situation K, which is presupposed in Scriven's example, there is no property which is assumed to be relevant for paresis. Therefore, the expected probability of p developing paresis is very low, say 0.001. If R is the narrowest reference class to which p belongs and F is the class of individuals suffering from paresis, then, in K, we have (using "P" instead of "P_W"): $P(F,R)=P(F)=0.001$; and according to $D4$ the belief value $B(Fp)$ is the same, *i.e.*, 0.001. Let S be the class of persons suffering from syphilis. K is enriched to $K_{T\cup C}$ where T contains at least the two statistical sentences $p(F,S)=0.1$ and $p(F,\bar{S})=0$, and $C=\{Sp\}$. For the situation $K_{T\cup C}$ containing this additional knowledge we obtain, by $D4$, that $B_{T\cup C}(Fp)=0.1$. According to $D5$ we are given a case of single case substantiation. This is so, although the final estimated probability is still low, because its value is one hundred times higher than it was in K. It is exactly this type of case which I, following Gärdenfors, intuitively singled out as exhibiting a considerable decrease of surprise value (or, equivalently, a corresponding increase of belief value). If, finally, one gets the additional information that Fp, then one may say that the whole constitutes an explanation of Fp, taking "explanation" in the first little story sense of this word. Intuition tells us that this is in accordance with common usage. Again, we must not forget that, after all, not-Fp would have been expected to a much higher degree than what really took place.

Why are we not able to reconstruct Hempel's counter-example along similar lines? The reason is very simple: we must assume that everyone knows that an individual p who has won the first prize has bought a ticket, and that this is the only thing relevant to his or her winning the prize. If F denotes the unit set of persons who win the first prize, S denotes the set of persons who have bought a ticket and R has a similar meaning as before, then in the underlying knowledge situation is $P(F)=P(F,R)=P(F,R\cap S)$ and therefore $B(Fp)=B(Fp,Sp)$. In other words, no increase in the belief value takes place.[11]

The reader should not forget what I have been claiming here and what I have not. It was *not* my intention to replace a given explication of *statistical explanation of events* by a better one, not even in the more specific sense in which I spoke of *single case substantiation* instead of *explanation.* All I wanted to do was to show how the reconstruction of knowledge situations along Gärdenfors's lines is much superior to that in terms of accepted sentences, and to illustrate how smoothly the difficulties, connected with an example which had turned out recalcitrant to prior analysis, disappear if we are willing to accept the change in reconstruction. As small as the change may appear in retrospect, it has one important new implication: the "prediction" is done neither in terms of subjective probabilities alone nor in terms of objective probabilities alone but in terms of *probability mixtures.*

VI. SUMMARY

Let me, for the moment, use the expression "*systematization*" in the wide Hempelian sense which covers all kinds of explanations, predictions, retrodictions etc. Then I can formulate my *philosophical* endeavour by saying that I was concerned with the question whether and in what sense the so-called statistical systematizations of singular events are of an *argumentative* type, as claimed by C.G. Hempel and P. Gärdenfors, or whether they are of a different, *non-argumentative* structure, as claimed by W. Salmon.

It turned out that there *is* an argumentative type of systematization, but it is restricted to the *predictive* use (as a special case of statistical inference): the "explanandum" is "rational statistical prediction" and *not* "statistical explanation." Furthermore, it turned out that there *is* a non-argumentative type of systematization which *may be called* explanation. But explanations of this kind can *never* be used as answers of why-questions since the word "explanation" is used in contexts similar to those in which we speak of explaining the working of a more or less complicated mechanism, and actually meaning by this *giving a detailed analysis of the mechanism.*

In the "little story sense of explanation" the word "explanation" may be used in *both* cases, but only at the cost of explicitly accepting an ambiguity in the expressions "to explain" and "explanation." For, in the first case, to explain something means to predict it rationally, by using the techniques of statistical inference, and, *in addition*, to observe later that it actually happened. What is "explained" here in a *strict* sense is the *rationality* of the prediction and not the actual outcome. In the second case, it means to analyze the working of the statistical mechanism and, *in addition*, to determine the actual outcome. However, it must not be overlooked that, in *both* cases, the course of events might have been different. This is of importance particularly in the first case, because an outcome not in accordance with the prediction is neither a falsification of the laws used nor of the rationality of the prediction. The only thing we can say in such a situation is that something happened which we could *not rationally have expected* to happen.

Perhaps it is the thesis of the structural symmetry of explanations and predictions,[12] tacitly accepted by many philosophers, which prevented a clear insight into this matter. Once the proper insight is gained, we may, in order to overcome the dilemma (A), substitute the one or the other story for "explanation," as the case may be. But if one is in search for something which is a statistical explanation of an event and at the same time a potential prediction of this event and *vice versa*, and if, in addition, one requires that this explanation must always be an answer to a why-question, then that search is hopeless. For *there is no such thing as a statistical explanation that simultaneously satisfies all of those conditions.*

If we look at all the possible cases considered, we find them overlapping and partially excluding each other, and yet all of them are still in accordance with our ordinary use of "explanation." Then, it seems to me, that in contrast to the view of Suppes's we ought to be careful and *very explicit*. For undifferentiated use of "statistical explanation," which *looks like* good scientific practise, will almost inevitably degenerate into "cocktail-party talk." This occurs even if we restrict ourselves to the cases where events and not *kinds* of events are "explained."

NOTES

1. My claim that explanation is an achievement concept may be mistaken. It may be that the English word "explanation" is not used as an achievement concept in all contexts. But the German counterpart of "explanation," *viz.* "Erklärung", is certainly always used in this way.

2. In [15], §7.

3. See, for example, Stegmüller [12], §5 and 6, in particular 6.e: "Denken in Likelihoods and Bayesianismus."

4. I take "prediction" to be the paradigm for the class of cases in which, intuitively speaking, we give "reasons" but not "causes." The so-called retrodictions, *e.g.*, are nothing but other members of this class.

5. [1], p. 9.

6. *Ibid.*

7. Moreover, the intuition is wrong also in the deterministic case. Suppose a person p suffers from a serious illness such that one can be sure that p will die within two or three weeks after t_0. But an explanation of the death of p which appeals to the fact of his illness may be quite wrong. p might have been shot dead one hour after t_0. This "screening off" of potential causes by rival ones inevitably leads us to the problem of causality which, however, I do not want to discuss here.

8. See Niiniluoto [5], p. 350.

9. See Scriven [9].

10. See Hempel [2], p. 369.

11. Gärdenfors mentions in [1], p. 13, that a similar analysis of the paresis example has been given by B. Hansson. Unfortunately, I have not had access to Hansson's paper.

12. For an extensive critical discussion of this thesis, see Stegmüller [10], Ch. II.

REFERENCES

1. Gärdenfors, P., *A Pragmatic Theory of Explanation*, Lund, Sweden, Working Paper No. 19, 1976.
2. Hempel, C.G., *Aspects of Scientific Explanation*, New York/London, 1965.
3. Hempel, C.G., "Maximal Specificity and Lawlikeness in Probabilistic Explanation," in: *Philos. of Sci.* 35 (1968), pp. 116–133.
4. Jeffrey, R.C., "Statistical Explanation versus Statistical Inference," in: Salmon, W.C. (ed.), *Statistical Explanation and Statistical Relevance*, pp. 19–28.
5. Niiniluoto, I., "Inductive Explanation, Propensity and Action," in: Manninen, J. and R. Tuomela (eds.), *Essays on Explanation and Understanding*, pp. 335–368.
6. Salmon, W.C. (ed.), *Statistical Explanation and Statistical Relevance*, Pittsburgh 1971.
7. Salmon, W.C., "Statistical Explanation," in: Salmon, W.C. (ed.), *Statistical Explanation and Statistical Relevance*, pp. 29–87, "Postscript 1971", pp. 105–110.
8. Salmon, W.C., "A Third Dogma of Empiricism," in: Butts, R.E. and J. Hintikka (eds.), *Basic Problems in Methodology and Linguistics*, Part Three of the Proceedings of the Fifth International Congress of Logic, Methodology and Philosophy of Science, Dordrecht 1977, pp. 149–166.
9. Scriven, M., "Explanation and Prediction in Evolutionary Theory," *Science* CXXX, (1959) pp. 477–482.
10. Stegmüller, W., *Wissenschaftliche Erklärung und Begründung*, Berlin-Heidelberg-New York, 2nd ed. 1974.
11. Stegmüller, W., *Personelle und statistische Wahrscheinlichkeit*, first half-volume: *Personelle Wahrscheinlichkeit und rationale Entscheidung*, Berlin-Heidelberg-New York 1973.
12. Stegmüller, W., *Personelle und Statistische Wahrscheinlichkeit*, second half-volume: *Statistisches Schließen – Statistische Begründung – Statistische Analyse*, Berlin-Heidelberg-New York 1973.
13. Stegmüller, W., "The Problem of Induction: Hume's Challenge and the Contemporary Answers," in: Stegmüller, W., *Collected Papers on Epistemology, Philosophy of Science and History of Philosophy*, Vol. II, Synthese Library 1977, pp. 68–136.
14. Stegmüller, W., "Carnap's Normative Theory of Inductive Probability," in: Stegmüller, W., *Collected Papers on Epistemology, Philosophy of Science and History of Philosophy*, Vol. II, Synthese Library 1977, pp. 137–149.
15. Stegmüller, W., *The Structuralist View of Theories. A Possible Analogue of the Bourbaki-Programme in Physical Science*, New York 1979.
16. Suppes, P., "Probabilistic Inference and the Concept of Total Evidence," in: Hintikka, J. and P. Suppes (eds.), *Aspects of Inductive Logic*, Amsterdam 1966, pp. 49–65.
17. Suppes, P., "Popper's Analysis of Probability in Quantum Mechanics," in: Schilpp, P.A. (ed.), *The Philosophy of Karl Popper*, Vol. I, La Salle, Ill., 1974, pp. 760–774.
18. Suppes, P., "New Foundations of Objective Probability: Axioms for Propensities," in: Suppes, P., Henkin, L., Moisil, Gr.C. and A. Joja (eds.), *Logic, Methodology and Philosophy of Science*, Vol. IV, Amsterdam 1973, pp. 515–529.
19. Wright, G.H.v., *Explanation and Understanding*, London 1971, pp. 13–15.

RODERICK M. CHISHOLM

(*Brown University*)

EPISTEMIC REASONING AND THE LOGIC OF EPISTEMIC CONCEPTS

1. THE BASIC SYSTEM

There are two approaches to the logic of epistemic concepts. One is to view it as a field of inquiry analogous to alethic logic or the logic of necessity. The other is to view it as a branch of the logic of preferability. In the present paper, I will take the second approach, making use of the following concepts: (a) *epistemic preferability*; (b) *de re* necessity; (c) *obtaining*, or *taking place*; and (d) *acceptance*, or *belief*.

We begin, then, with the locution, "*p* is epistemically preferable to *q* for *S* at *t*" — or "*p* is more reasonable than *q* for *S* at *t*" — where the expressions occupying the place of "*p*" and "*q*" are terms referring to states of affairs and where "*S*" and "*t*," respectively, refer to a particular person and to a particular time.

In the following statement of the principles of epistemic logic, the expressions "*Bh*," "*Wh*," "*P*," and "*S*" may be taken to abbreviate respectively, "accepting *h*," "withholding *h*," and "is more reasonable for *S* at *t* than," and "is the same in epistemic value for *S* at *t* as."[1] "Withholding *h*" may be taken, in turn, to abbreviate, "Not accepting *h* and not accepting not-*h*."

(A1) For every *e*, *h*, and *i*, if it is false that *Be P Bh*, and if it is false that *Bh P Bi*, then it is false that *Be P Bi*.

(A2) For every *h* and *i*, if *Bh P Bi*, then it is false that *Bi P Bh*.

(A3) For every *h*, if it is false that *Wh P Bh*, then *Bh P B*$\neg$*h*.

(A4) For every *h* and *i*, *Bh P Bi*, if and only if, *B*$\neg$*i P b*$\neg$*h*.

(A5) For every *h* and *i*, *Wh S Wi*, if and only if, either *Bh S Bi* or *B*$\neg$*h S Bi*.

(A6) For every *h* and *i*, if *Bi P Bh* and *Bi P B*$\neg$*h*, then *Wh P Wi*.

(A7) *h* is identical with $\neg\neg$*h*.

G.H. von Wright (ed.), Logic and Philosophy, 71–78.

The final principle, which may be thought of as belonging to the general theory of states of affairs rather than to epistemic logic, enables us to deduce that withholding *h* is the same as withholding the negation of *h*.

In order to explicate the basic concepts of the theory of epistemic preferability, we should consider what is involved in asking, for any given proposition, any given subject and any given time, which is epistemically preferable: accepting the proposition, accepting the negation of the proposition, or withholding the proposition. In this way we may now explicate a number of fundamental epistemic concepts. For simplicity, the temporal reference is omitted from most definitions.

D1.1 *h* is *beyond reasonable doubt* for *S* =Df. Accepting *h* is more reasonable for *S* than is withholding *h*.

D1.2 *h* has *some presumption in its favour* for *S* =Df. Accepting *h* is more reasonable for *S* than accepting not-*h*.

D1.3 *h* is *acceptable* for *S* =Df. Withholding *h* is not more reasonable for *S* than accepting *h*.

D1.4 *h* is *certain* for *S* =Df. *h* is beyond reasonable doubt for *S*, and there is no *i* such that accepting *i* is more reasonable for *S* than accepting *h*.

D1.5 *h* is *counterbalanced* for *S* =Df. Accepting *h* is not more reasonable for *S* than accepting not-*h*, and accepting not-*h* is not more reasonable for *S* than accepting *h*.

Among the consequences of our principles and definitions are the following:

For every state of affairs *h*, either *h* is unacceptable for *S* at *t* or not-*h* is unacceptable for *S* at *t*. Hence no state of affairs is "indifferent" in the sense of being such that both it and its negation are acceptable. (But a state of affairs may be "indifferent" in the sense of being counterbalanced.)

For every state of affairs *h*, *h* falls into one and only one of the following seven categories, for *S* at *t*: (1) *h* is beyond reasonable doubt; (2) *h* is acceptable but not beyond reasonable doubt; (3) *h* has some presumption in its favor but is not acceptable; (4) *h* is counterbalanced; (5) not-*h* has some presumption in its favor but is not acceptable; (6) not-*h* is acceptable but not beyond reasonable doubt; and (7) not-*h* is beyond reasonable doubt.

The principles also imply that withholding a state of affairs that is counterbalanced is epistemically preferable to withholding a state of affairs that is not counterbalanced.

2. APPLICATION TO EPISTEMIC FOUNDATIONS

We now add the concept of *de re* necessity to the foregoing system. This is the concept expressible in the locution "x is necessarily such that it is F." It is here restricted to states of affairs.

By thus extending the system, we can characterize: (a) the concept of an epistemic *basis*, or *foundation*; (b) those relations in virtue of which one state of affairs may be said to confer epistemic status upon another; and (c) the analogy between practical reasoning and epistemic reasoning.

The definitions that immediately follow are schematic; the letter "F" is replacable by any predicative expression.

In the first definition we introduce "self-presentation" as an absolute concept, a concept holding eternally of states of affairs. It is assumed that some states of affairs are necessarily such that they cannot obtain without being the object of someone's certainty.

D2.1 The state of affairs, something being F, is self-presenting =Df. The state of affairs, something being F, is necessarily such that, for every x and for any time t, if x is F at t, then something being F is certain for S at t.

D2.2 It is self-presenting for S at t that he then has the property of being F =Df. S has the property of being F at t; and the state of affairs, something being F, is self-presenting.

D2.3 The state of affairs, something being F, is self-presenting for S at t =Df. It is self-presenting for S at t that he then has the property of being F.

We may now characterize what is directly evident *a posteriori*:

D2.4 h is *directly evident a posteriori* for S =Df. h is logically contingent; and there is an e such that (i) e is self-presenting for S and (ii) necessarily, whoever accepts e accepts h.

In one of its traditional senses the word "axiom" is used to refer to a proposition which is necessarily such that, if one understands it, then one sees that it is true. I believe that the sense of this conception is captured by the following definition:

D2.5 h is *an axiom* =Df. h is necessarily such that (i) it obtains and (ii) for every S, if S accepts h, then h is certain for S.

An alternative to this definition could be obtained by substituting "entertains" for "accepts."

D2.6 h is directly evident *a priori* for S =Df. (i) h is an axiom and (ii) *S accepts h.*

D2.7 h is directly evident for S =Df. Either h is directly evident *a posteriori* for S, or h is directly evident *a priori* for S.

3. THE EVIDENT

Although we have introduced the concept of the directly evident, we have not yet characterized the more general concept of the evident.

An adequate characterization of the evident would be one enabling us to say that the evident is that which distinguishes knowledge from true belief which is not knowledge. What is known should have an epistemic status higher than that of being beyond reasonable doubt; yet it need not be certain. And so we must specify an epistemic category which falls between that which is certain and that which is beyond reasonable doubt.[2]

To single out this category, we will turn to what might be called "conferring relations," those relations in virtue of which one state of affairs may be said to confer positive epistemic status upon another. One such relation may be expressed by saying "e tends to confirm h" and defined as follows:

D3.1 e tends to confirm h =Df. e is necessarily such that, for every subject x, if e is beyond reasonable doubt for x and if everything that is beyond reasonable doubt for x is logically implied by e, then h has some presumption in its favor for x.

D3.2 h is evident for S =Df. h is beyond reasonable doubt for S; and no conjunction of states of affairs tends to confirm not-h.

The following principle may be thought as belonging to epistemology rather than to epistemic logic: there are states of affairs which are such that they may be evident for a subject without being certain for that subject.

4. EPISTEMIC RELATIONS AND PROBABILITY

The definitions that follow pertain to epistemic relations that may obtain between states of affairs. They will enable us to distinguish four different epistemic uses of the expression "probability."

Our first definition may be said to give us the *logical* sense of probability, that sense of probability which may be construed as a relation holding eternally between propositions or states of affairs.

D4.1 *e* makes *h prima facie* probable =Df. *e* is necessarily such that for every *S*, if *e* is evident for *S* and if everything that is evident for *S* is entailed by *e*, then *h* has some presumption in its favor for *S*.

The definiens may also be read as "*e* confirms *h*" or "*h* is more probable than not in relation to *e*."

A second use of "probability" is that wherein a state of affairs may be said to be *prima facie* probable for a given subject *S*:

D4.2 *h* is *prima facie* probable for *S* at *T*=Df. There is an *e* such that (i) *e* is evident for *S* and (ii) *e* makes *h* probable.

A third use of "probability" is that wherein a state of affairs is said to make another state of affairs probable *for* a given subject. We thus take note of the way in which the logical concept of probability may function epistemically for a particular subject:

D4.3 *e* makes *h* probable for *S* =Df. *e* is evident for *S*; *e* makes *h prima facie* probable; and there is no *i* such that (a) *i* is evident for *S* and (b) *e* & *i* does not make *h prima facie* probable.

The second clause of the foregoing definition could also be put by saying: "There is no *i* such that (a) *i* is evident for *S* and (b) *i* overrides the confirmation that *e* provides for *h*." The relevant sense of "override" may be defined as follows:

D4.4 *i overrides* the confirmation that *e* provides for *h* =Df. *e* makes *h prima facie* probable; and the conjunction, *e* and *i*, does not make *h prima facie* probable.

The following states of affairs illustrate the concept expressed by "*i* overrides the confirmation that *e* provides for *h*":

(e) Most *F*'s are *G*'s, and *a* is *F*

(h) *a* is *G*

(i) Most *F*'s which are also *H*'s are not *G*'s, and *a* is *H*

It should be noted that the overriding of a confirmation may itself be overridden.

We now introduce a fourth use of "probability." This is what is sometimes called the *absolute* concept of probability for a given subject at a given time.[3]

D4.5 *e* is more probable than not for *S* =Df. There is something that makes *h* probable for *S*.

We may assume that, if a state of affairs is thus more probable than not for *S*, then it has some presumption in its favor for *S*.

It is sometimes said that "there is no such thing as *the* probability of a state of affairs, or of a statement asserting one, but different probabilities on different data,"[4] But the foregoing distinction suggests that there *is* such a thing as *the* probability of a state of affairs for a given person at a certain time, its absolute probability for that person at that time.

We note, in passing, two "conferring relations" that are stronger than those previously considered:

D4.6 *e* is a *basis* of *h* for =Df. *e* is directly evident for *S*; and necessarily, if *e* is directly evident for *S*, then *h* is evident for *S*.

D4.7 *e confers evidence* upon *h* for *S* =Df. *e* is evident for *S*; and every *b* such that *b* is a basis of *e* for *S* is a basis of *h* for *S*.

5. THE ANALOGY BETWEEN PRACTICAL AND EPISTEMIC REASONING

The concept of confirmation may be said to function in our epistemic reasoning in a way analogous to that in which the concept of requirement functions in our practical reasoning.[5] Consider the following practical arguments:

(A) (1) *p* occurs;
(2) *p* requires that *S* performs *A*;

therefore

(3) *S* has a *prima facie* duty to perform *A*.

The conclusion of this argument follows from the premises if we define "*S* has a *prima facie* duty to perform *A*" by saying "something occurs which requires *S* to perform *A*."

Now it is quite possible for a person to have conflicting *prima facie* duties. Thus the premises of argument (A) are consistent with those of the following argument (B):

(B) (1) q occurs:
(2) q requires that S not perform A;

therefore

(3) S has a *prima facie* duty not to perform A.

From the fact, then, that a person has a *prima facie* duty to perform a certain action A, it does not follow that he has an absolute duty to perform that action A. He has an absolute duty only if he has a *prima facie* duty which has not been overridden. In other words:

(C) (1) There occurs an x which is such that x requires that S perform A;

(2) There occurs no y such that the conjunction of x and y does not require that S perform A;

therefore

(3) S has an absolute duty to perform A.

The application of probability theory is similar. It can easily happen that there are two evident propositions p and r which are such that (i) p makes p *prima facie* probable and (ii) p & r does not make q *prima facie* probable. In such a case, as we have noted, r may be said to *override* the confirmation that p provides for q. The relation between *prima facie* duty and absolute duty, then, has its analogue in the relation between what we have called the *prima facie* probability of a state of affairs and the *absolute probability* of that state of affairs.

If, then, we are applying the theory of probability in a particular case, we may have arguments analogous to the three practical arguments above:

(A) (1) p is evident for S at t;
(2) p confirms q;

therefore

(3) q is *prima facie* probable for S at t.

The expression "p confirms q," in premise (2), is an alternative reading for "p makes q *prima facie* probable." The conclusion of the argument follows from the premises since we have defined "q is *prima facie* probable" by saying "something that is evident for S at t makes q *prima facie* probable for S at t." Now the premises of argument (A) are consistent with those of the following argument (B):

(B) (1) r is evident for S at t;
(2) r confirms not-q;

therefore

(3) Not-q is *prima facie* probable for S at t.

Hence one and the same state of affairs may be confirmed as well as disconfirmed for a given subject at a given time. From the fact that a state of affairs is thus *prima facie* probable for a given subject or a given time, it does not follow that the state of affairs is *absolutely* probable for that subject at that time. A state of affairs is absolutely probable only if its *prima facie* probability has not been overridden. In other words:

(C) (1) There is an x such that x is evident for S at t and x confirms q;
(2) There is no y such that y is evident for S at t and the conjunction of x and y does not confirm q;

therefore

(3) q is absolutely probable for S at t.

Hence epistemic reasoning is similar in essential respects to practical reasoning.

NOTES

1. Versions of the first five axioms were used in "A System of Epistemic Logic" by Roderick M. Chisholm and Robert Keim, *Ratio* 15 (1973), pp. 99–115.

2. The following definition was proposed in "A System of Epistemic Logic": an *evident* proposition is a proposition which is beyond reasonable doubt and which is such that believing it is epistemically preferable to withholding any proposition that is counterbalanced. (Cf. *ibid.*, p. 114.)

3. See Bernard Bolzano, *Theory of Science*, ed. by Rolf George, Basil Blackwell, Oxford 1972, pp. 359–365. Cf. G.E. Moore, *Commonplace Book*, Allen & Unwin, London 1962, pp. 401–402.

4. C.I. Lewis, *An Analysis of Knowledge and Valuation*, Open Court Publishing Co., La Salle Ill. 1946, p. 267.

5. I have discussed the concept of requirement in detail in "Practical Reason and the Logic of Requirement," in: *Practical Reason*, ed. by Stephen Körner, Basil Blackwell, Oxford 1964, pp. 40–53.

EVANDRO AGAZZI

(*University of Genova*)

ON CERTAINTY, EVIDENCE AND PROBABILITY

Professor Chisholm's paper is such an excellent essay in philosophical analysis, that one cannot do anything but admire the fine work he has performed in bringing forth so many detailed and relevant concepts. In such a situation, the task of a commentator, if he is expected to do something more interesting than expressing his general approval and appreciation, might perhaps be that of trying to propose a few clarifications or improvements on some particular points and this is what I am actually going to do. Let me only say that, owing to shortage of time, I was unable to read other papers by Chisholm which are referred to in his present one, so that I cannot exclude that answers to some of my questions might be found there. It seems useful, however, to raise these questions at least in order to make his present paper more self-contained.

I

A first need of clarification may be found in his definition D1.4 (p. 72) of the concept of *certain*, which states:

D1.4 *h* is *certain* for *S* =Df. *h* is beyond reasonable doubt for *S*, and there is no *i* such that accepting *i* is more reasonable for *S* than accepting *h*.

The second member in the conjunction of the definiens: "there is no *i* such that accepting *i* is more reasonable for *S* than accepting *h*" might leave open some undesired possibilities because of the "absolute" character of this definition (and of the others in this section as well). For instance, that the electron in a hydrogen atom has a negative charge may be considered certain (both in the intuitive and in Chisholm's sense) in physics. On the other hand, the statement that $2+2=4$ might be said to be more reasonable to accept for a subject at a given time *t*, because it is much simpler in its mathematical foundation than the first is within physics. If *S*

G.H. von Wright (ed.), Logic and Philosophy, 79–84.

were to make a very engaging bet, *e.g.*, it is quite likely that he would find it more reasonable to rely upon the truth of "$2+2=4$" than upon the truth of "the electron has a negative charge," not because he has any *reasons* for questioning the certainty of the physical statement, but because the certainty of the mathematical statement at issue is, so to speak, more immediate, and this fact seems to play some role and to deserve some consideration in an epistemic context. The way to avoid this difficulty might be found in a relativization of the concepts at issue to a given universe of discourse or to some existing evidence, in order to give to the "more reasonable" condition some flavour of conflicting or rival instances coming on the stage.

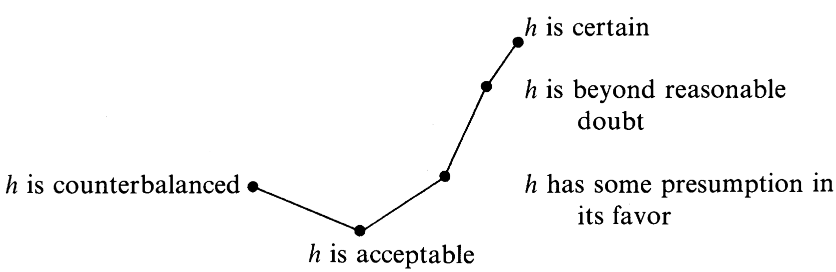
But even so one can find similar situations leading to difficulties. Let us consider a formal system with a set of postulates P which are accepted as certain by a given subject S. Let now T be a theorem provable in the given formal system and actually proved in it, at time t. One must surely maintain that T is certain as well, in any reasonable intuitive meaning of "certain." Yet it might not fully satisfy Chisholm's second requirement, for it seems hardly deniable that, if S were to express his epistemic preference between T and one of the postulates of P which is necessarily involved in the proof, he could not help finding it more *reasonable* to accept the postulate. The "reason" would be that, owing to the peculiar way of establishing T (*i.e.*, by means of a formal proof), the certainty of P is a necessary prerequisite for the certainty of T, but not *vice versa*. Or, to put it differently: the "intrinsic possibilities" of being wrong are more numerous in the case of T, since they can be hidden in the complexity of the proof, even if it is factually correct and unobjectionable.

There might be a philosophical conception supporting Chisholm's definition, namely the claim that certainty admits of no graduation, that it is a matter of all-or-nothing, so that all certain propositions are on the same footing. I shall not deny the philosophical attractiveness of this thesis. Still it might prove too strong a claim, while the above arguments of mine suggest a more realist conception which admits different "degrees of certainty." They are different from the obviously existing "degrees of certitude" (which are of a psychological nature) in that they correspond to some actual "degrees of complexity" of the way of reaching certainty which can in many cases be made explicit.

A way which seems rather close to Chisholm's intention and which is neutral as far as the graduability of certainty is concerned, would be: to relativize the notion of certainty to every universe of discourse or available evidence and then to introduce a new notion of "rival" propositions (e.g., resorting to some suitable idea of mutual incompatibility). In this case, definition D1.4 might assume the following unobjectionable form:

D1.4′ *h* is *certain* for *S* =Df. *h* is beyond reasonable doubt for *S*, and there is no rival *i* such that accepting *i* is more reasonable for *S* than accepting *h*.

Another point on which I feel some perplexities is the identification of Chisholm's seven categories for states of affaires (p. 72). He says: "(2) *h* is acceptable but not beyond reasonable doubt; (3) *h* has some presumption in its favour but is not acceptable" (analogous statements are put forth for not-*h*). My impression is that this partition is a more or less formal consequence of Chisholm's previous statements, rather than of a conceptual analysis of the very notions involved. I should venture therefore to suggest a different partition, which seems to me more faithful to the actual use of the notions of something "being acceptable," "being beyond reasonable doubt" and "having some presumption in its favor." If my proposal should be found convincing, some reformulation of Chisholm's axioms and definitions would be needed. This I am not going to explore, however. My point is that, according to his definitions, to be acceptable seems to express a weaker condition for *h*, than to have some presumption in its favor. As a matter of fact, definition D1.3 seems to indicate that *h* being acceptable corresponds to the minimal requirement of being "free from objections," so that it would not prove reasonable to "suspend judgement" on it, or to "withhold *h*," as Chisholm puts it. At this stage, no comparison of *h* with not-*h* needs be involved or, to put it differently, *h* simply appears as *prima facie* tenable: a quite common situation which occurs whenever one makes an assumption believing it to be true (notice that for Chisholm acceptance means belief, as he says at page 71). After this first step, one starts inquiring whether there are good reasons for defending *h*, and if this turns out to be the case, one can conclude either that the reasons in favor of *h* are as strong as those in favor of not-*h* or that one of the two propositions relies upon stronger reasons than the other. If this happens for *h*, we can say that *h* has some presumption in its favor, according to D1.2. If our analysis is correct, we could represent the increasing strength of the epistemic requirements for *h* by means of the following tree:



As a consequence, Chisholm's categories (2) and (3) should be modified as follows: (2′) h has some presumption in its favor but is not beyond reasonable doubt; (3′) h is acceptable but has no presumption in its favor. Analogous modifications should affect also (5) and (6) concerning not-h.

Another fact which is not explained in the paper is why the cases of h being certain (or not-h being certain) are not considered or admitted. Does it mean that when h (or not-h) is beyond reasonable doubt it is also certain? And why should this always be the case?

II

I now come to Chisholm's set of definitions concerning "self-presentation." Definition D2.1, which introduces the "absolute" formulation of this concept, should, in my opinion, be slightly improved by adding an existential quantifier and receive the following formulation (in which the proposed addition is indicated within brackets):

D2.1′ The state of affairs, something being F, is self-presenting =Df. The state of affairs, something being F, is necessarily such that, for every x and for any time t, if x is F at t, then (there is an S such that) something being F is certain for S at t.

This improvement seems to me necessary in order to avoid relativization of self-presentation to a particular S, which seems contrary to Chisholm's aim. On the other hand, if one accepts my modification, definition D2.2 becomes problematic, for it says:

D2.2′ It is self-presenting for S at t that he has the property of being F =Df. S has the property of being F at t; and the state of affairs, something being F, is self-presenting.

This definition is all-right if self-presentation is relativized to S, but it is not if we adopt my proposed modified version of D2.1 which preserves absoluteness. For this would imply that D2.2 becomes explicitly:

D2.2′ It is self-presenting for S at t that he then has the property of being F =Df. S has the property of being F at t; and the state of affairs, something being F, is self-presenting *for S at t*.

This move, however, seems to be forbidden by Chisholm's definition D2.3, which introduces explicitly the notion of "self-presentation for S" in the following form:

D2.3 The state of affairs, something being F, is self-presenting for S at t =Df. It is self-presenting for S at t that he then has the property of being F.

If one accepts this definition, however, D2.2′ becomes circular. My impression is that in D2.3 "it" should be put in place of "he", or that, more clearly, the definition should be restated as follows:

D2.3′ The state of affairs, some x being F, is self-presenting for S at t =Df. It is self-presenting for S at t that x then has the property of being F.

If one accepts D2.3′ and puts it before D2.2′, everything seems to be suitably settled.

I do not know whether I misunderstood Chisholm completely, but it seems to me that my remarks are well taken unless he tacitly maintains that "self-presentation" is somehow a kind of "self-consciousness" of S. In this case he could keep his definitions as they stand, but at the price of relativizing self-presentation to a particular S. From the context of his discourse I have the impression that neither of these is his intention.

III

As for Section 3 on "The Evident," I fully agree with Chisholm's effort to find an epistemic category falling between that which is certain and that which is beyond reasonable doubt. I should only express a terminological preference not to use the word "evidence" for this intermediate status, but to use, for instance, some expression like "fully reliable." One reason for that is that Chisholm himself has so beautifully characterized the notion of "directly evident" in Section 2, that it does not seem advisable to weaken it without strong reasons. I should prefer simply to call "evident" what he calls "directly evident" and try to find, as I said, a new term for the intermediate case which he is correctly pointing at. Another reason for so doing is mostly historical: in the Western philosophical tradition, the notion of evidence has almost invariably been bound to that of certainty (as also Chisholm's "direct evidence" does), both in the case of "empirical evidence" (comparable with his *a posteriori* one) and in the case of "logical evidence" (comparable with his *a priori* one). It is commendable to preserve this feature especially because then evidence could continue to play its essential role of providing a "foundation" for certainty. This means, in particular, that one of the possible grounds for it being "more reasonable for S to accept h rather than not-h" (see definition D1.4 of "certain") could be the presence of some evidence in favor of *h*. I am quite aware that this might not be fully harmonizable with Chisholm's present treatment of the concept of evidence, in which certainty is a defining condition for *a priori* evidence, but it seems to me that this point could be slightly modified without altering the substance of his views. The third reason for not recommending a separation between evidence and certainty is that it might unconsciously suggest a wrong tenet, which is moreover semantic and not epistemic. This tenet could be expressed by saying that an evident proposition might not be true, *i.e.*, not correspond to what is actually the case. I think that Chisholm does not support this "gnosiological dogma," but then he should find here a good reason for not introducing such a separation between evidence and certainty.

IV

An analogous terminological preference I should like to express regarding the notion of "probability" used in Section 4. A word like "plausibility" would better fit in, in my opinion, with the very fine analysis Chisholm develops in this section. As a matter of fact, the concept he defines through several ingenious steps remains a qualitative one and does not lend itself to any measurement, contrary to what is expected to be the case with probability proper. Notice that this holds true also for so-called "subjective probability" of Keynes, De Finetti and others, so that Chisholm's concept does not cope with this notion of probability either, which is only superficially epistemic in character. A word like "plausibility," on the contrary, seems to me to agree much better with the epistemic intention of Chisholm's discourse and, in particular, to avoid the somewhat awkward notion of an "absolute probability" to which no precise *value* can be assigned. To call it, *e.g.*, "plain plausibility" might perhaps sound quite appropriate or, if one prefers to stress the strength of this plausibility, one might use the terminology "reliably plausible" for the proposition *h* to which Chisholm applies the qualification "absolutely probable." This terminological choice would also avoid the striking impression one gets when he speaks of "probability theory" (p. 77) to denote his topic, while this locution is currently used to designate a precise and sophisticated mathematical discipline which seems to have not much in common with the kind of discourse he is developing here.

But these are, after all, only minor remarks and I should rather want to finish by stressing the special elegance of the analogy between epistemic and practical reasoning, which concludes Professor Chisholm's very interesting paper.

INDEX OF NAMES