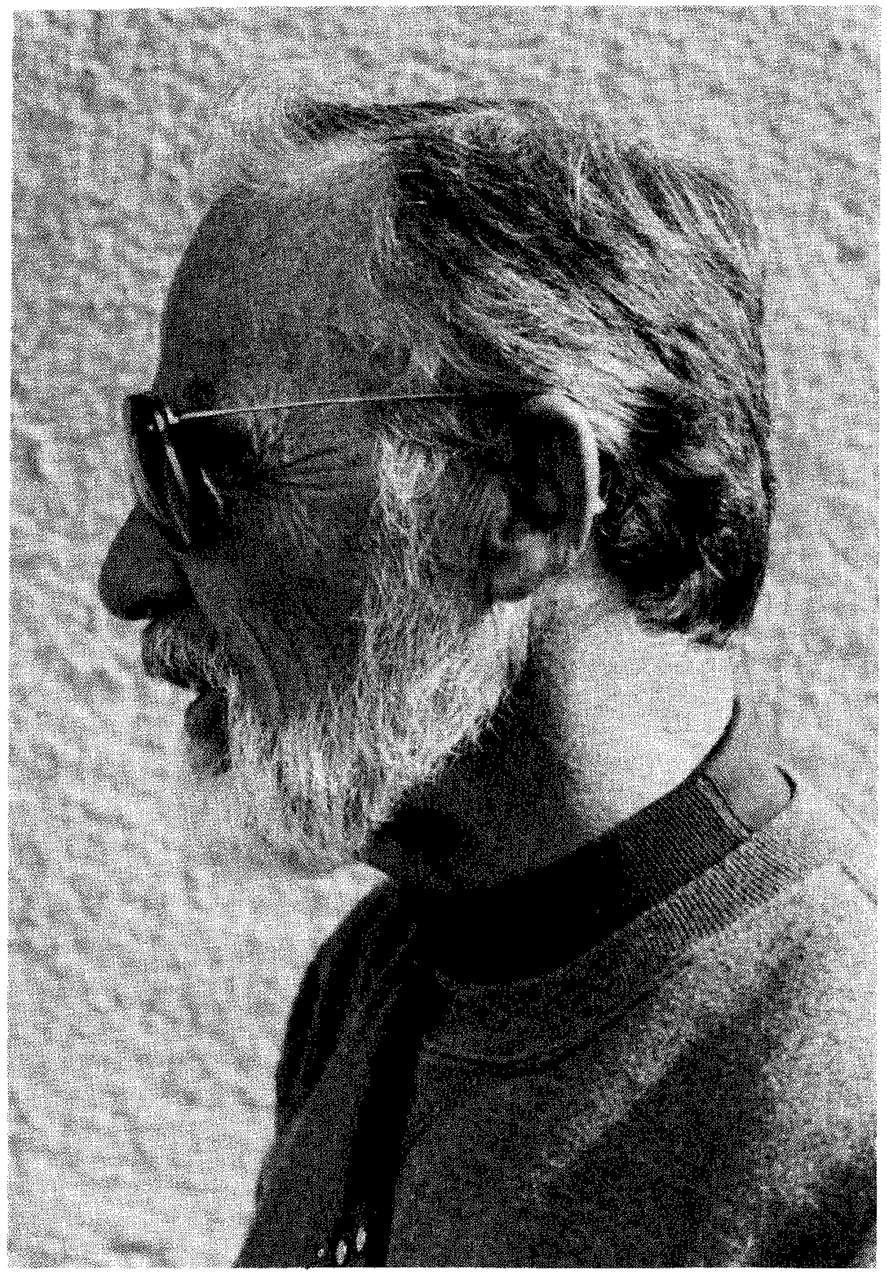
THE UNIVERSITY OF WESTERN ONTARIO SERIES
IN PHILOSOPHY OF SCIENCE

A SERIES OF BOOKS
IN PHILOSOPHY OF SCIENCE, METHODOLOGY, EPISTEMOLOGY,
LOGIC, HISTORY OF SCIENCE, AND RELATED FIELDS

VOLUME 49



KAREL LAMBERT

EXISTENCE AND EXPLANATION

Essays presented in Honor of Karel Lambert

Edited by

WOLFGANG SPOHN
University of Bielefeld

BAS C. VAN FRAASSEN
Princeton University

and

BRIAN SKYRMS
University of California at Irvine

KLUWER ACADEMIC PUBLISHERS
DORDRECHT / BOSTON / LONDON

Library of Congress Cataloging-in-Publication Data

Existence and explanation : essays presented in honor of Karel Lambert
/ edited by Wolfgang Spohn, Bas C. van Fraassen, Brian Skyrms.
p. cm. -- (The University of Western Ontario series in
philosophy of science ; v. 49)
Includes bibliographical references and index.
ISBN 0-7923-1252-X (HB : acid free paper)
1. Science--Philosophy. 2. Philosophy, Modern. 3. Lambert,
Karel, 1928- . I. Lambert, Karel, 1928- . II. Spohn,
Wolfgang. III. Van Fraassen, Bas C., 1941- . IV. Skyrms, Brian.
V. Series.
Q175.3.E95 1991
501--dc20 91-15399

ISBN 0-7923-1252-X

Published by Kluwer Academic Publishers,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Kluwer Academic Publishers incorporates the publishing programmes
of D. Reidel, Martinus Nijhoff, Dr W. Junk and MTP Press.

Sold and distributed in the U.S.A. and Canada
by Kluwer Academic Publishers,
101 Philip Drive, Norwell, MA 02061, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers Group,
P.O. Box 322, 3300 AH Dordrecht, The Netherlands.

printed on acid free paper

Printed in The Netherlands

To Joe

TABLE OF CONTENTS

WOLFGANG SPOHN

A REASON FOR EXPLANATION: EXPLANATIONS PROVIDE STABLE REASONS

1. INTRODUCTION*

Why ask 'Why?' Whence our drive for explanation? This is a bewildering question because it is hard to see what an answer might look like. I well remember having learnt in undergraduate courses that explanation is the supreme goal of science. So who would dare ask for more? Some fortunately did.[1] One prominent answer is that (scientific) explanation yields (scientific) understanding; and surely, we want to understand things. It is this answer which this paper is about.

When I first heard of this answer from Karel Lambert as being seriously discussed, it struck me as utterly tautological; and when arguing against it in Lambert (1988, 1991) he seemed to argue for a contradiction. However, there is one, and only one, way of rendering this answer sensible and sensibly doubtable: namely by giving independent characterizations of '(scientific) explanation' and of '(scientific) understanding' and checking how they relate. This is what Lambert (1988, 1991) did, thus recovering the full worth of the answer. But it is not what is usually done. Quite often the correctness of the answer has been presupposed, and ideas about what understanding might consist in have been built into the characterization of explanation.[2] But then the answer helps only to explicate, not to justify explanation.

Lambert (1991) concluded that the fact that an answer to a why-question is an explanation is neither sufficient nor necessary for it to yield understanding. I want to advance an argument in favour of the contrary conclusion. It needs a double stage-setting (sections 2 and 4) and has two steps (section 3 and 5). Section 2 mainly presents a general theory of non-probabilistic induction. This is the basis for section 3: a partial account of deterministic causation (which copies the probabilistic account I have given in (1983) and (1990a)) and a straightforward extension thereof to a partial account of causal explanation. Section 4 works up to some coherentistic principles in terms of the given theory of induction which involve what I shall call ultimately stable reasons. The notion of an ultimately stable reason cannot pretend

Wolfgang Spohn et al. (eds.), Existence and Explanation, 165–196.

to catch much of the rich notion of understanding; but, as section 5 explains, it fits well the characterizations of understanding which have been given in this context and may thus serve as a substitute.[3] Section 5 finally proves a formal equivalence of causal explanations and ultimately stable reasons under some restrictions which require several comments. Since the epistemological relevance of ultimately stable reasons unfolds in a coherentistic picture of truth, this equivalence construes the search for explanation as the search for truth.

2. INDUCTION AND CAUSATION

David Hume was the first to argue for an essential connection between induction and causation, so forcefully in fact that it has not ceased since to be in the focus of philosophical discussion. Indeed, for Hume induction and causation were virtually the same:

Although Hume himself struggled with the characterization of belief — believing, he said, is having ideas accompanied by a peculiar feeling of vivacity and firmness[4] —, he has an elaborate theory of belief formation. Impressions as the most lively and forceful of all perceptions are the paradigms and the basis of belief; all other empirical beliefs are gained from them by inductive extension. How? Hume held that induction proceeds just by inferring causes from effects and vice versa, i.e. via causal inferences which sufficiently, though not completely, transfer the impressions' vivacity and firmness so characteristic of belief.[5] The realm of empirical belief therefore consists of nothing but causal inferences from impressions (and their recollections).

Induction thus seems to reduce to causation. But one may as well view the matter the other way around. Hume defines causation, taken as what he calls a natural relation, as precedence, contiguity and association, i.e. transfer of liveliness and firmness, the marks of belief.[6] Thus, if A precedes B and is contiguous to it, A is a cause of B if and only if B may be inductively inferred from A. This shows that induction and causation are in effect interdefinable for Hume.[7]

The imperfections of Hume's account are well known. It is certainly wrong to say that A is a cause of B if and only if B may be inductively inferred from A, even if A and B satisfy the other conditions; symptoms of later events are clear counter-examples. And if one gives up this equivalence, it is doubtful that causal inferences exhaust inductive inferences. However, I believe that such imperfections do not defeat

Hume's fundamental insight into the essential connection between induction and causation; the task is to get it straight.

Since the ways of induction seem multifarious, it is implausible that induction should be definable in causal terms. Thus one part of this task, the one discussed in the rest of this section, must be to provide a general account of induction independent of causation. The other part, taken up in the next section, is then to say how causation relates to this account.

Concerning the first part, the first point to note is that induction and belief revision are one and the same topic: The input of an inductive scheme consists of all the information directly received, and it tells what to believe according to the input. The input of a scheme of belief revision consists of an old epistemic state and a new piece of information, and it yields a new epistemic state. Thus, a scheme of belief revision may be immediately inferred from an inductive scheme, and the latter follows from the former plus an initial epistemic state to start from. This congruence may not always have been clear because induction and revision have met different interests. Belief revisionists explicitly searched only for rationality constraints on belief, whereas the longer-standing discussion of induction tended to search for *the* correct inductive scheme, thereby presupposing, or perhaps only hoping, that there is just one such scheme possibly even independent of the initial epistemic state. History taught us, I think, that this presupposition, or hope, is misguided.[8] Therefore, the two fields have merged by now, and general accounts of induction may best be found by looking at accounts of belief revision.

Within the representation of epistemic states as (subjective) probability measures, belief revision is a rich and lively topic.[9] However, Instead want to turn to a much less familiar representation of epistemic states. One essential weakness of the probabilistic representation is that it can hardly account for *plain* belief which simply holds propositions to be true or false or neither; this is the moral of the well-known lottery paradox.[10] Therefore I dismiss probability because I want to focus on plain belief — for several reasons: First, it is of intrinsic interest to examine the structure of inductive schemes for plain belief. Secondly, if induction and causation are indeed essentially connected, then, presumably, subjective probabilities are related to probabilistic causation, whereas (sufficient and/or necessary) deterministic causation relates to plain belief; and it is the latter kind of causation I am concerned with.

Thirdly, the probabilistic counterparts of some of the assertions in the final section hold only under more restrictive conditions. Fourthly, and perhaps most importantly, subjective probabilities cannot be true or false; truth attaches only to plain belief; thus an important part of my argument will only work for plain belief.

Strangely enough, induction and revision with respect to plain belief is a much more experimental and less established field. Shackle's functions of potential surprise, Rescher's plausibility indexing, and Cohen's inductive probability[11] are pioneering contributions, and revision of plain belief has been most thoroughly studied by Gärdenfors and his coauthors.[12] In (1988) and (1990b) I have proposed a slight variant of these epistemic representations which has the advantage that it allows of generally and iteratedly applicable revision rules and thus in effect of a full account of induction for plain belief. Its basic concept is easily introduced:

Throughout, Ω is to be a set of *possible worlds* (as philosophers say without necessarily being so serious about it as is, e.g., David Lewis) or a sample space (as probability theorists prefer to say), i.e. just an exhaustive set of mutually exclusive possibilities; elements of Ω will be denoted by α, ν, ω, etc. As usual, *propositions* are represented by subsets of Ω, denoted by A, B, C, D, E, etc. The basic concept is then given by

DEFINITION 1. κ is a *natural conditional function*[13] (a *NCF*) iff κ is a function from Ω into the set of natural numbers[14] such that $\kappa^{-1}(0) \neq \emptyset$. A NCF κ is extended to propositions by defining $\kappa(A) = \min \{\kappa(\omega) | \omega \in A\}$ for each $A \neq \emptyset$ and $\kappa(\emptyset) = \infty$.[15]

A NCF κ is to be interpreted as a *grading of disbelief*. If $\kappa(\omega) = 0$, then ω is not disbelieved, i.e. ω might be the actual world according to κ. Because not every world can be denied to be the actual one, Definition 1 requires that $\kappa(\omega) = 0$ for some $\omega \in \Omega$. If $\kappa(\omega) = n > 0$, then ω is disbelieved with degree n. A proposition is then assigned the minimal degree of disbelief of its members.[16] Thus, if $\kappa(A) = n > 0$, then A is disbelieved with degree n. And if $\kappa(A) = 0$, then A is not disbelieved, i.e. A might be true according to κ. $\kappa(A) = 0$ does not mean that A is believed according to κ. Belief in A is rather expressed by disbelief in $\overline{A}$[17], i.e. by $\kappa(\overline{A}) > 0$, i.e. $\kappa^{-1}(0) \subseteq A$. Thus, $\kappa^{-1}(0)$ determines what is plainly believed according to κ.

Two simple properties of NCFs should be noted: *the law of negation* that for each proposition A either $\kappa(A) = 0$ or $\kappa(\overline{A}) = 0$ or both, and *the law of disjunction* that for all propositions A and B, $\kappa(A \cup B) = \min(\kappa(A), \kappa(B))$.

According to a NCF κ, propositions are believed in various degrees. It is useful to explicitly introduce the function expressing these degrees, because it is more vivid than the above disblief talk:[18]

DEFINITION 2. β is *the belief function associated with* the NCF κ iff β is a function from the power set of Ω into the set of integers extended by $+\infty$ and $-\infty$ such that $\beta(A) = \kappa(\overline{A}) - \kappa(A)$.[19] β is a *belief function* iff it is associated with some NCF.

Thus, $\beta(\overline{A}) = -\beta(A)$, and A is believed true or false or neither according to β (or κ) depending on whether $\beta(A) > 0$ or < 0 or $= 0$.[20]

So far, the various degrees of belief did not really play a theoretical role. But they are crucial for an account of belief revision, the central notion of which is this:

DEFINITION 3. Let κ be a NCF and A a non-empty proposition. Then *the A-part of* κ is the function $\kappa(\cdot | A)$ defined on A by $\kappa(\omega | A) = \kappa(\omega) - \kappa(A)$ for each $\omega \in A$. Again, this function is extended to all propositions by setting $\kappa(B | A) = \min \{\kappa(\omega | A) | \omega \in A \cap B\} = \kappa(A \cap B) - \kappa(A)$ for each $B \subseteq \Omega$. Finally, if β is the belief function associated with κ, we define, as in Definition 2, $\beta(B | A) = \kappa(\overline{B} | A) - \kappa(B | A)$.

Definition 3 immediately implies *the law of conjunction* that $\kappa(A \cap B) = \kappa(A) + \kappa(B | A)$ for all propositions A and B with $A \neq \emptyset$, and *the law of disjunctive conditions* that $\kappa(C | A \cup B)$ is between $\kappa(C | A)$ and $\kappa(C | B)$.[21]

The A-part $\kappa(\cdot | A)$ of κ can be viewed as a NCF with respect to the restricted possibility space A and thus as a grading of disbelief *conditional on* A. Accordingly, $\beta(\cdot | A)$ expresses degrees of belief conditional on A.

It is obvious that a NCF κ is uniquely determined by its A-part $\kappa(\cdot | A)$, its $\overline{A}$-part $\kappa(\cdot | \overline{A})$, and the degree $\beta(A)$ of belief in A. This suggests a simple model of belief revision for NCFs. If a piece of

information consists only in the proposition A, then it is plausible to assume that only the old degree $\beta(A)$ of belief in A gets changed to some new degree $\beta'(A) = n$, whereas the A-part and the $\overline{A}$-part of the old NCF κ are left unchanged; n, $\kappa(\cdot|A)$, and $\kappa(\cdot|\overline{A})$ then determine a new NCF κ', the revision of the old κ by that information.[22] There are also more complicated models in which the information need not concern a single proposition. These suggestions indicate that NCFs indeed allow for a theory of revision and induction for plain belief. But there is no need to further develop the theory of NCFs.[23] The sequel requires mainly an intuitive grasp of the notions introduced in Definition 1—3.

A first useful application of these notions is the concept of a reason. Being a reason is always relative to an epistemic background, and given such a background a reason strengthens the belief in, or, in other words, is positively relevant to, what it is a reason for. This intuition can be immediately translated into formal terms:

DEFINITION 4. Let β be the belief function associated with the NCF κ, and A, B, and C three propositions. Then A is *a reason for B relative to β* (or κ) iff $\beta(B|A) > \beta(B|\overline{A})$. And A is *a reason for B conditional on C relative to β* (or κ) iff $\beta(B|A \cap C) > \beta(B|\overline{A} \cap C)$.

Note that, according to this definition, the relation of being a reason is *symmetric, but not transitive*, in analogy to probabilistic positive relevance, but in sharp contrast to the narrower relation of being a deductive reason (which is just set inclusion between contingent propositions[24]). Note also that, according to this definition, being a reason does not presuppose that the reason is actually given, i.e. believed; on the contrary, whether A is a reason for B relative to β is independent of the degree $\beta(A)$ of belief in A.

Since the value 0 has the special role of a dividing line between belief and disbelief, different kinds of reasons can be distinguished:

DEFINITION 5.

$$A \text{ is a } \left\{\begin{array}{l} \textit{additional} \\ \textit{sufficient} \\ \textit{necessary} \\ \textit{weak} \end{array}\right\} \textit{ reason for } B \textit{ relative to } \beta \text{ (or } \kappa\text{) iff } \left\{\begin{array}{l} \beta(B|A) > \beta(B|\overline{A}) > 0 \\ \beta(B|A) > 0 \geqslant \beta(B|\overline{A}) \\ \beta(B|A) \geqslant 0 > \beta(B|\overline{A}) \\ 0 > \beta(B|A) > \beta(B|\overline{A}) \end{array}\right\}.$$

Conditional reasons of the various kinds are defined similarly. If A is a reason for B, it belongs at least to one of these four kinds; and there is just one way of belonging to several of these kinds, namely by being a necessary and sufficient reason. Though the emphasis will be on sufficient and on necessary reasons, the two other kinds, which do not show up in plain belief and are therefore usually neglected, well deserve to be allowed for by Definition 5.

3. CAUSATION AND EXPLANATION

Ultimately, this section will arrive at a (partial) explication of causal explanation. But this will be only a small step beyond saying how causation is related to the general account of induction for plain belief just formally introduced. So, let me turn to the latter task.

A is a cause of B, as a first approximation, iff A and B both obtain, A precedes B, and under the obtaining circumstances A raises the epistemic or metaphysical rank of B. Most people can agree on this vague characterization, the disagreement is only about how to precisely understand it. It's thus a good start; four points call for comment.

(1) 'A and B obtain': The precise nature of the causal relata A and B is a serious problem beyond the scope of this paper. I just take them to be propositions; since I did not say much about what propositions are except that they are subsets of Ω, this can hardly be wrong. No one doubts that the causal relata have to obtain, to be facts, or to be actual. This entails that causation is world-relative, i.e. that the explicandum rather is 'A is a cause of B in ω'. In the given framework, the condition that A and B obtain in ω is simply expressed by the clause that $\omega \in A \cap B$.

(2) 'A precedes B': Some think that backwards causation should not be excluded by definition, and some more think that at least instantaneous causation should be allowed. I am not sure. But since this is not my present concern, I will just stick to the temporal precedence of the cause.

But so far, there is no time in possible worlds; they need a bit more structure: Let I be a non-empty set of factors or variables; each variable $i \in I$ is associated with a set Ω_i containing at least two members; Ω_i is the set of values i may take. The set Ω of possible worlds is then represented as the Cartesian product of all the $\Omega_i (i \in I)$. Thus, each ω

$\in \Omega$ is a course of events, a function assigning to each variable $i \in I$ the value $\omega(i)$ which i takes in the possible world ω.

I shall call I a *frame* and say that Ω and its elements are *generated* by the frame I and that a NCF on Ω and its associated belief function are *for* the frame I. Already here it is clear, and to be emphasized, that the explication of causation given below will be frame-relative. This is unavoidable, if the explication is to be expressed in formally well-defined terms. Though this frame-relativity seems to me to be natural, one may find it awkward that A is a cause of B within one frame, but not within another. From the present position this relativity can only be overcome by moving into a fictitious universal frame I^* which is not further extensible. Since we shall have occasion in the next section to indulge in that fiction, we may at present be content with this relativity.

Time may now simply be represented by a weak order $\leqslant$, i.e. a transitive and connected relation, on the frame I (since metric properties of time are irrelevant); $<$ denotes the corresponding irreflexive order on I; and for $j \in I$, $I_{<j}$ is to be the set $\{k \in I \mid k < j\}$. I shall neglect the complications of continuous time and assume that time is discrete.

Time should be associated not only with variables, but, if possible, also with propositions. Therefore we define a proposition A to be a *J-measurable* or, in short, a *J-proposition* for a set $J \subseteq I$ of variables iff for all $v, \omega \in \Omega$ agreeing on J, i.e. with $v(i) = \omega(i)$ for all $i \in J$, $v \in A$ iff $\omega \in A$. Maximally specific J-propositions will be called J-states; thus, A is a *J-state* iff A is J-proposition and any two $v, \omega \in A$ agree on J. In particular, ${}^{\omega}J$ is to denote the J-state $\{v \in \Omega \mid v$ agrees with ω on $J\}$.

There are many contingent propositions which are about a single variable, and the temporal order of variables is easily carried over to them. Indeed, I see no loss at all in restricting causes and effects to be such, so to speak, logically simple propositions which are about one variable. The condition that A precedes B will thus be expressed by requiring that there are $i, j \in I$ with $i < j$ such that A is a contingent i-proposition and B a contingent j-proposition.

(3) 'A raises the epistemic or metaphysical rank of B': The clumsiness and obscurity of this phrase is due to its intended generality. That A raises the rank of B simply means that the rank of B given A is higher than given $\overline{A}$; thus the phrase makes sense only if conditional ranks are defined. With respect to probabilistic causation, these ranks

are probabilities, of course; and they are epistemic or metaphysical ranks according to whether probabilities are interpreted subjectively as degrees of belief or objectively as chances. With respect to deterministic causation, the phrase covers all kinds of approaches — regularity theories, counterfactual theories, analyses in terms of necessary and/or sufficient conditions (however these are understood in turn), etc. — which differ on the interpretation of metaphysical and epistemic ranks. What I shall propose is easily anticipated: I shall take ranks to be epistemic ranks as given by belief functions in the sense of Definition 2. Thus, this is the point where I, following Hume, trace the essential connection between induction and causation.

Why should one follow Hume and conceive causation as an idea of reflexion, as he calls it?[25] Why construe the apparently realistic notion of causation as essentially epistemically relativized? Why try to say not what causation *is*, but rather what the causal conception of a subject in a given epistemic state is? After all, Hume himself was not so unambiguous; his definition of causation as a philosophical relation is pure regularity theory void of any epistemic elements, and when stealing the realist's thunder in (1739), pp. 167—169, his emphasis is on that definition. I cannot do justice here to this profound problem, which even provoked Kant's so-called Copernican revolution; let me just mention my two main motives for taking Hume's side.

One reason is quite concrete. The literature is full of examples presenting problems to various explications of causation, and an explication of causation relative to belief functions is, I believe, more successful in coping with these problems than rival accounts. I will expand a bit on this claim after the formal explication.

The other reason is that there is not only a strong realistic intuition of causation, but also an urgent need for explaining the most prominent epistemological role of the notion of causation. If causation is epistemically relativized, this explanation ensues naturally. But without such a relativization, I do not know of a good explanation. If causation is conceived as a kind of physical ingredient of the world (say, energy transfer), the explanation would have to go like this: "There are a lot of people around, and I can't fail to notice them; therefore, people play an important role in my world picture. Similarly, there is a lot of causation around, and I can't fail to notice it; this explains the prominent epistemological role of the notion of causation." But that parallel sounds wrong to me; it underrates the peculiar epistemological importance of causa-

tion, which is different from that of people and other ubiquitous things. And if causation is conceived as a kind of structural component of the world (say, a deductive relation between laws of nature and singular facts, or a relation of counterfactuality, or a certain relation between universals), the explanation must be given in terms which cannot be accepted without further elucidation. Such terms may be lawlikeness for a regularity theory, similarity between possible words for the counterfactual account of Lewis (1973),[26] a theoretical relation of causal necessitation between universals for Tooley (1987), sect. 8.3.2, etc.; and I am not convinced that there are unproblematic ways of objectively understanding these terms.

Of course, the realistic intuition of causation should not be forgotten because of the epistemological concern. Hume did not forget it, as his two definitions of causation as a philosophical and a natural relation show, in which regularity is the objective counterpart to subjective association. Any more adequate implementation of Hume's general strategy has to make the same kind of move. In particular, it is incumbent on me to say under which conditions there is a kind of objective counterpart to NCFs or belief functions.[27] However, here I will be content with the primary, epistemically relativized explication.[28] These remarks may also make the above-mentioned frame-relativity more plausible.

(4) 'Under the obtaining circumstances': This phrase is also beset with difficulties. In particular, it seems that the relevant circumstances of A's causing B are all the other causes[29] of B which are not caused by A; and this renders the initial characterization of causation patently circular.[30] However, the circularity dissolves, if only A's being a *direct* cause of B is considered; there are then no intermediate causes, i.e. no causes of B which are caused by A, and thus the relevant circumstances may be conceived as consisting of all other causes of B. Moreover, it seems to do no harm when all the irrevelant circumstances are added, i.e. all the other facts preceding, but not causing B; and thereby the obtaining circumstances of A's directly causing B may be conceived as consisting of all the facts preceding B and differing from A. This is what I propose:

DEFINITION 6. Let $\omega \in \Omega$, $i,j \in I$, A be an i-proposition, and B a j-proposition. Then A is *a direct cause of B in ω relative to* the belief function β (or the associated NCF κ) iff $\omega \in A \cap B$, $i < j$,[31] and

$\beta(B|A \cap {}^{\omega}(I_{<j}\text{-}\{i\})) > \beta(B|\overline{A} \cap {}^{\omega}(I_{<j}\text{-}\{i\}))$, i.e. A is a reason for B conditional on ${}^{\omega}(I_{<j}\text{-}\{i\})$ relative to β. More specifically, A is *an additional, sufficient, necessary, or weak direct cause of B in ω* according to whether A is an additional, sufficient, necessary, or weak reason for B conditional on ${}^{\omega}(I_{<j}\text{-}\{i\})$.

In my (1980), pp. 79ff., and (1983), pp. 384ff., I have more fully argued that ${}^{\omega}(I_{<j}\text{-}\{i\})$, i.e. the state in ω of all the variables preceding the effect and differing from the cause is indeed the correct proposition to conditionalize on; that is, I have argued that whenever we base our judgment about the direct causal relation between A and B on fewer facts, it could be just the neglected facts which would change the judgment. This is confirmed by the more detailed investigation into the relevant circumstances of causal relations in sect. 4 of my (1990a).

I believe that causation in general should be defined as the transitive closure of direct causation, as seems quite natural and as many have assumed. A fully defense of this view, however, is a long story, parts of which I have told in my (1990a). For the present purpose, it suffices to consider only direct causation.

To make Definition 6 a bit more vivid, it may be helpful to briefly explain how it deals with three standard difficulties. The first is the problem of irrelevant law specialization introduced by Salmon (1970), pp. 177ff., which says that, according to the original Hempel-Oppenheim account, John's regularly taking birth control pills explains his not becoming pregnant. Regularity theories of causation are of course threatened by this problem, too. But there is no problem for Definition 6. Given John is a man (before the given time of his non-pregnancy), his regularly taking contraceptives (before that time) is just irrelevant to, and not a reason for, his non-pregnancy at that time, at least relative to our educated belief functions.

The second problem is the distinction between causes and symptoms which is a graver obstacle to regularity accounts of causation and explanation. Take, e.g., C. D. Broad's Manchester hooters and London workers discussed at length by Mackie (1974), pp. 81ff. Whenever the factory hooters in Manchester and London sound, which is the case every working day at 6 p.m., then the workers in Manchester and London leave their work shortly afterwards. But only the London and not the Manchester hooters have an impact on the London workers. Again, this case presents no problem for Definition 6. Unconditionally, the

proposition that the Manchester hooters sound (at a particular day) is relevant to the proposition that the London workers leave; but given that the London hooters sound (or do not sound), the former is just irrelevant to the latter. Again this is true relative to our normal belief functions. Of course, one may have a different belief function yielding also a conditional relevance; but then the sounding of the Manchester hooters is not treated as a mere symptom of the London workers' leaving.

The general scheme should be clear by now. NCFs and belief functions help us to notions of (conditional) relevance and irrelevance which are much more sensitive than the relevance notions provided by other approaches to deterministic causation and which behave almost the same as probabilistic relevance notions.[32] Thus, they enable us to copy the methods of dealing with these problems which have been so successfully developed for probabilistic causation.

A third problem further illustrating this scheme distinguishes Definition 6 not only from regularity theories, but also from counterfactual analyses of causation like Lewis (1973). It is the problem of (symmetric) causal overdetermination cruelly, but standardly, exemplified by the firing squad. Prima facie, cases of causal overdetermination are clearly possible. But as far as I see, they are a great mystery, if not an impossibility for all realistic accounts of causation; it seems that the only thing the realist can do is to explain them away: either by observing that one of the two causal chains from the two apparently overdetermining causes to the effect has not been completed so that one of the two causes is in fact preempted; or by observing that there is an intermediate event (a Bunzl event, as Lewis (1986), p. 208, calls it) which causes the effect and which is jointly caused by the two apparently overdetermining causes so that the two causes in fact jointly cause the effect.[33]

For Definition 6, however, there is no mystery at all. The two overdetermining causes may be simply conceived as additional causes; each of the two is an additional cause of the effect in the presence of the other one. The crucial difference is that additional causes cannot be defined within a counterfactual approach, let alone a regularity theory. Something true can counterfactually be still true or not true, but not more or less true. But something conditionally believed can be believed more or less firmly under different conditions.[34] Of course, I do not claim that this simple remark solves the problem of causal overdetermination. What it does is first to do justice to the prima facie existence

of causal overdetermination and secondly to locate the problem; it arises when we try to objectivize our epistemically relativized causal picture, because there is no realistic counterpart to additional causes as defined above.

So much for the partial account of causation we need. It is easily extended to a partial account of explanation. I shall not comment on explanation of laws and theories and on whether there is non-causal explanation.[35] But concerning causal explanation, it seems unassailable to say that getting an explanation for B is learning a cause of B and having an explanation for B is knowing a cause of B.[36] The problem is only that this statement is unhelpful as long as one does not have an account of causation or tries to explain causation by explaining explanation. But this is not our problem, and thus we may immediately turn this informal statement into a formal definition:

Knowing some fact to be a cause at least involves believing this fact to be a cause. And believing A to be a cause of B according to a NCF κ means believing the actual world to be among the worlds in which A is a cause of B relative to κ. Since only direct causes have been formally defined, this leads to

DEFINITION 7. Let i, j, A, and B be as in the previous definition. *A's range $C_{A,B}$ of directly causing B relative to* the NCF κ or its associated belief function β is defined as the $I_{<j}$-$\{i\}$-measurable set of all $\omega \in \Omega$ such that $\beta(B \mid A \cap {}^{\omega}(I_{<j}\text{-}\{i\})) > \beta(B \mid \bar{A} \cap {}^{\omega}(I_{<j}\text{-}\{i\}))$. Hence $A \cap B \cap C_{A,B}$ is the set of all $\omega \in \Omega$ such that A is a direct cause of B in ω relative to κ or β.[37] *A's range ${}^{s}C_{A,B}$ or ${}^{n}C_{A,B}$ of*, respectively, *sufficiently or necessarily directly causing B relative to* κ or β is defined accordingly. Then, *A causally explains B (as necessary, as possible) relative to* κ or β iff $\beta(A \cap B \cap C_{A,B}) > 0$ ($\beta(A \cap B \cap {}^{s}C_{A,B}) > 0$, $\beta(A \cap B \cap {}^{n}C_{A,B}) > 0$).[38]

The only deviation of Definition 7 from its informal statement is that knowledge of a cause has been weakened to belief in a cause. This corresponds to the old debate whether explanation requires true or only accepted antecedents. I think there are both usages; 'B is explained by A' may be factive or not according to whether it is taken as the passive of the apparently factive 'A explains B' or as an ellipsis of the apparently non-factive 'B is explained by A by some explainer'. Since I have always talked only of belief and not of knowledge, I settle for the

weaker version. I do not see that our topic is seriously affected by this issue. In particular, I do not see that this issue drives a wedge between explanation and understanding, as Lambert (1988), pp. 308—310, and (1991), pp. 138f., argues. Understanding as well can be taken factively or non-factively, and it seems only fair that, when assessing the relation between explanation and understanding, only the corresponding interpretations are compared.

4. REASON AND TRUTH

The first task of giving a partial account of explanation need not be developed further.[39] The next task is to give an independent account of understanding or, rather, of some not too bad substitute thereof. I approach this task by discussing three principles of increasing strength which I take to be basic principles of coherence, believability, and truth.

Let's start with a simple question: If B is a contingent proposition, is there a reason for B? Trivially, yes. There always are deductive reasons; each non-empty subset of B is a sufficient, and each non-tautological superset a necessary, deductive reason for B. So, the question should rather be whether there is an *inductive*, i.e. non-deductive reason for B. Or, put in another way, if B is a contingent j-proposition for some $j \in I$, is there a I-$\{j\}$-proposition which is a reason and thus an inductive reason for or, for that matter, against B?[40] Not necessarily, of course. There may be variables which are independent of all other variables in the given frame I relative to the given belief function β; and since I may be an arbitrary collection of variables, such counterinstances ensue naturally.

Consider now an extension I' of the frame I and an extension β' of β for I'. There are many such extensions of β, and, trivially, there exists an extension of β according to which there is a I'-$\{j\}$-measurable reason for B. Thus, we should, more precisely, consider *the* extension β' of β as determined by some unspecified epistemic subject with a belief function covering also I'-propositions. Is there a I'-$\{j\}$-measurable reason for B according to β'? Again, not necessarily. The case of I' is not different from the case of I.

But now consider all extensions I' of I and the appertaining belief functions β' which are within the subject's range. Should then an inductive reason for or against B come to the fore? Once more, not necessarily; but that would be a grave matter. It would mean that the subject could not learn anything whatsoever about B; wherever it looked, it could not find the slightest hint concerning B; B would be outside its world of experience, outside its bounds of sense.

It may of course happen that a proposition is beyond a subject's present grasp. This may change; an individual's inductive scheme as well as that of a society or of mankind keeps evolving.[41] It may even be that a proposition is forever beyond the grasp of an individual or of actual mankind. But these are all accidental limitations. My real concern is the status of the j-proposition B with respect to all possible extensions of the frame I whatsoever and the appertaining belief functions to which a subject would extend its belief function β, if it came to consider these extensions of I. Is it still conceivable that in this sense no extension of I contains an inductive reason for B? Now, finally, it seems plausible to say no. Otherwise, there would be no way whatsoever to reason or to learn anything about B, not because of accidental limitations, but due to the inherent structure of the all-inclusive inductive scheme underlying all these extensions of β; B would be literally senseless, unreasonable.

I have referred to all possible extensions of some initial frame and inductive scheme. But it is simpler to refer instead to the universal frame I^* comprising all variables whatsoever, to the set Ω^* of possible worlds generated by I^*, and to a universal belief function β^* for I^*. It may seem earthlier to talk only of extensions. But the set of all extensions is not earthlier than its union; both are philosophical fiction. Talking of I^*, Ω^*, and β^* is just much less clumsy than quantifying over extensions.

I^*, Ω^*, and β^* are what, in a loosened usage of Kantian and Peircean terms, has been called regulative ideas, ideal limits of inquiry, etc. The question whether one can legitimately and sensibly appeal to such limit concepts is certainly pressing. Here I just follow all those who do so. And I take it that, insofar our epistemic activities may at all be described by frames and belief functions, we conceive these activities as embeddable into the universal frame I^* and a universal belief function β^* and that we consider this embeddability as a fundamental requirement of consistency.[42]

What we have arrived at, then, is a first plausible principle of coherence:

(PCo1) For any $j \in I^*$ and any contingent j-proposition B there is a I^*-$\{j\}$-measurable reason for B relative to β^*.

Pco1 may be taken as a condition on β^*, on how β^* has to connect

propositions. But it may also be conceived as a condition on I^* (and the generated Ω^*) saying that no logically simple proposition exists unless appropriately connected by β^*. The best is to view PCo1 as what it is, as a condition simultaneously on I^* *and* β^*.

PCo1 is, of course, akin to the positivists' verifiability principle and other criteria of empirical significance. But PCo1 is a weak version, because it requires at best confirmability and not verifiability and because it does not refer to a directly verifiable basis, to evidential certainty, and the like. And PCo1 is unambiguous about the nature of the required ability of confirmation. This ability is not to be taken as restricted by our sensory outfit; PCo1 does not refer to any specific senses. It is not restricted by limited computing capacities; β^* will not be computationally manageable, anyway. It is not restricted by our spatiotemporally and causally limited access to facts. This ability is constituted exclusively by the inherent structure of the limiting inductive scheme and thus of the actual inductive schemes approaching it.

Given the above explication of direct causation, PCo1 is, by the way, tantamount to the following weak principle of causality:

(PCa1) For any $j \in I^*$ and any contingent j-proposition B there is a direct cause or a direct effect of B in some world $\omega \in \Omega^*$ relative to β^*.

At least the equivalence of PCo1 and PCa1 holds, if I^* is linearly and discretely ordered by time.[43] Note also that the reference to I^* and β^* eliminates the frame-relativity of that explication, but not its epistemological involvement.

PCo1 is symmetric with respect to positive and negative relevance; whenever a proposition is a reason for B, its negation is a reason against B. This symmetry will break in the next step when we consider *true* propositions; truth is biased towards positive relevance:

We have first to introduce another limit concept: the actual world taken not as a spatiotemporally maximally inclusive thing, but as everything that is the case. We naturally assume that among all the possible worlds in Ω^* exactly one is actual; let's call it α^*. Thus, a proposition A is true (absolutely, not relative to a model or a world) iff A is true in α^*, i.e. iff $\alpha^* \in A$.

The question now is this: Suppose that the contingent j-proposition B is true. PCo1 asserts that there are I^*-$\{j\}$-measurable reasons for B. But will there be a true reason among them? Let's look at the question

in a more earthly setting of a small frame I, the small actual world α (which is the restriction of α^* to I), and a subject's belief function β for I. Within this setting, the answer may certainly be no. If so, however, the truth would be undetectable, unbelievable for the subject within this setting. If it believed only truths, it would have no reason for believing B; and if it has any reason for believing B, then only by believing some I-$\{j\}$-propositions which are false. This situation is not critical by itself, but it again becomes more and more critical when it does not change as larger and larger extensions of I are considered. And relative to I^*, α^*, and β^*, finally, this situation seems absurd; all true evidence which could conceivably be brought to bear on B would univocally speak against B and for $\overline{B}$, though B is true and $\overline{B}$ false. Thus it seems plausible to answer the initial question in the affirmative.

This can be stated as a second principle of coherence:

(PCo2) For any $j \in I^*$ and any contingent j-proposition B with $\alpha^* \in B$ there is a I^*-$\{j\}$-measurable proposition A with $\alpha^* \in A$ which is a reason for B relative to β^*.

Briefly: for each singular truth there is a true inductive reason. Of course, PCo2 implies PCo1.

In Peirce-Putnamian terms one might say that PCo2 is part of the assertion that the epistemically ideal theory cannot be false. The ideal theory has, of course, recourse to all true evidence; and in a case violating PCo2 the ideal theory would have to falsely affirm $\overline{B}$ on the basis of that evidence and the universal inductive scheme β^*. PCo2 prevents this and thus captures at least one aspect of Putnam's internal realism.[44]

Indeed, PCo2 fits well under the heading 'coherence theory of truth'. The theoretical standing of the coherence theory is not exactly brilliant, because of difficulties in saying precisely what coherence is. Explications in deductive terms, say as consistency or deducibility, were precise, but unprofitable; and other, more interesting explications were always vague. A noticeable exception is Rescher (1973) and (1985); but I find his underlying theory of plausibility indexing not fully satisfying. Here, coherence is construed as inductive coherence as constituted by positive relevance relative to a belief function. PCo2 is thus one way of saying that truth must cohere. Of course, a workable theory of induction or belief revision for plain belief is vital to this construal.

PCo2 does certainly not yield a definition of truth. For propositions,

being true is defined as having α^* as a member; and for sentences, Tarski's truth definition may need an underpinning by a theory of reference, as called for by Field (1972), but as a definition it does not need a coherentist supplement. PCo2 also does not yield a criterion of truth; it is of little help in determining the truth of B because it is kind of circular in requiring true reasons for B and because it does not tell what to do in the case of conflicting reasons. In fact, PCo2 is not a condition on truth alone; it must again be viewed as a condition on I^*, α^*, *and* β^*, on how truth and reason relate in the universal frame.

There is also a principle of causality associated with PCo2:

(PCa2) For any $j \in I^*$ and any contingent j-proposition B with $\alpha^* \in B$ there is a direct cause or a direct effect of B in α^* relative to β^*.

Briefly: each singular fact has a direct cause or a direct effect in the actual world. This principle of causality is, of course, much stronger and much more interesting than PCa1. PCa2 is even stronger than PCo2; the former implies the latter, but not vice versa.[45] It would be nice to find a plausible principle of coherence entailing PCa2; so far I have not succeeded.

There are, however, plausible strengthenings of PCo2. One of them is my next goal.

PCo2 asserts that a true I^*-$\{j\}$-measureable reason A may be found for the contingent true j-proposition B. Now imagine that a piece C of true information is received and that A is then no longer a reason for B, i.e. A is not a reason for B conditional on C. This is not impossible; if A is positively relevant to B given one condition, A may be positively, negatively, or not relevant to B given another condition. And it is not excluded by PCo2. But this seems an implausible way to satisfy the plausible PCo2.

This opens up a new kind of question: How does the relevance of some truth to B evolve in the infinite process of acquiring more and more true information? Formally, everything is possible. The relevance may (a) vacillate for some (or no) time and then forever stay on the positive side, or (b) vacillate for some (or no) time and then forever stay on the non-positive side, or (c) vacillate forever. A truth of kind (b) is a very casual kind of reason for B, if at all, and one of kind (c) an odd and deeply undecided kind.

Is it conceivable that all true reasons for B one finds after some true information or other turn out to be of these unreliable kinds (b) and (c)? Formally, there are again three ways how this might happen. First, it might be that true reasons for B run out after sufficient true information. This case definitely violates the basic idea of PCo2 that in the limit all truth must be believable. Secondly, it might be that at infinitely many stages of the acquisition of true information there are true reasons for B and at infinitely many other stages there are no true reasons for B. This case again violates the basic idea. As often as one gains confidence in B, one loses it; one can never hold it fast. Thirdly, it might be that after some true information there always are true reasons for B, though different ones at each subsequent stage of the process. This case seems to be compatible with the basic idea, but it is still strange. Each time when asked why one believes B one has to withdraw the previous answer and to give another one; and this continues forever. This does not seem to be an acceptable process of truth tracking.

I therefore conclude that there should be at least one reason for B of the reliable kind (a); I shall call it an *ultimately stable* reason. This is the key concept of the following considerations; it is more precisely defined thus:

DEFINITION 8. Let $\omega \in \Omega$ and $A, B, C \subseteq \Omega$. Then A is a *ω-stable (sufficient, necessary) reason for B within C relative to* a belief function β for I (or the associated NCF κ) iff $\omega \in A \cap B$, $\omega \in C$, $A \cap C \neq \emptyset \neq \bar{A} \cap C$, and A is a (sufficient, necessary) reason for B relative to β conditional on each $D \subseteq C$ with $\omega \in D$ and $A \cap D \neq \emptyset \neq \bar{A} \cap D$. A is an *ultimately ω-stable (sufficient, necessary) reason for B relative to β* iff A is so within some condition. The set of all $\omega \in \Omega$ such that A is an ultimately ω-stable (sufficient, necessary) reason for B is called *A's range of being an ultimately stable (sufficient, necessary) reason for B* and denoted by $S_{A,B}({}^{s}S_{A,B}, {}^{n}S_{A,B})$.

Note that the truth of A and B in ω is made a defining characteristic of A's being an ultimately stable reason for B. Note also that, if A is an ultimately ω-stable reason for B, so is B for A.

In these terms, then, I have just argued for a third principle of coherence:

(PCo3) For any $j \in I^*$ and any contingent j-proposition B with $\alpha^* \in B$ there is a I^*-$\{j\}$-measurable, ultimately α^*-stable reason for B relative to β^*.

Briefly: for each singular truth there is an ultimately stable inductive reason. If there are reasons with stronger than ultimate α^*-stability, say, with α^*-stability within Ω, all the better. But such stronger forms of stability do not seem to be required in PCo3 on coherentistic grounds. Still, PCo3 implies PCo2.[46]

5. EXPLANATIONS AND STABLE REASONS

Now I can finally offer my substitute for (scientific) understanding: it is knowing ultimately stable reasons. I do not want to defend this as an explication of the complex notion of understanding. But what has been said in this context about understanding is captured fairly well by my proposal; and knowing ultimately stable reasons is epistemologically significant in its own right. Let me explain.

What is meant by knowing an ultimately stable reason A for B? Not only that one knows A and A is in fact an ultimately stable reason for B, but also that one knows A to be so, i.e. that one knows A's range $S_{A,B}$ of being an ultimately stable reason for B to obtain. As in the case of explanation, there is a factive and a non-factive understanding of understanding, and as in the former case I deal only with the latter, in order to be able to confine myself to belief and to be silent about knowledge. To be precise, then, A's being believed to be an ultimately stable reason for B relative to β simply comes to $\beta(S_{A,B}) > 0$.

The significance of believing in ultimately stable reasons is this: When one believes A and B to be true, one thinks that A and B are part of, fit into, α^* in some way or other. But one may do so as a mere recorder of facts without any understanding of what is going on, without any grasp of *how* A and B fit into α^*. And one may, adhering to PCo3, simply proclaim that it should be possible to find an ultimately α^*-stable reason for B. When one believes $S_{A,B}$ to be true, however, one does not only believe A and B, and one does not merely postulate an ultimately α^*-stable reason for B. Rather, one thinks to know a particular one, namely A. And one has a partial grasp of how A and B fit into α^*, namely as one element of coherence, as one coherent link among many others which have to exist in α^*. Thus, for the believer of $S_{A,B}$ A and B better qualify as part of the final truth than for the believer of A and B alone.

How else is understanding characterized? Lambert (1991), p. 129, says that "the metaphor of 'fitting into', and its stylisic variants such as

'incorporated into' or 'integrated into', seem especially germane vis à vis scientific understanding as it relates to scientific explanation" and quotes a number of important authors using this metaphor. For him, then, "the sense of scientific understanding relevant to scientific explanation may be characterized as an answer to the question 'How does state-of-affairs S fit into theory T?'" (p. 130), where, as he goes on to explain, "fit into" may mean various things.

Similarly, Friedman (1974) and Kitcher (1981), again adducing a number of witnesses, take unification as the key concept. Friedman (1974), p. 15, explicitly claims:

> . . . this is the essence of scientific explanation — science increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given. A world with fewer independent phenomena is, other things being equal, more comprehensible than one with more.

And he goes on to say more precisely how he understands independence or independent acceptability.

These seem to be appropriate ways of talking also about ultimately stable reasons; indeed, I myself slipped into these ways three paragraphs ago. Of course, the metaphors take on different senses with the different authors. But this is not an unhappy homonymy which hides incomparable interests. On the contrary, I think, there is one common idea which is vague and allows of various explications, and there is less a disagreement about its content, but rather a common need in surveying this idea and tracing fruitful explications. Here, in any case, fit and unification, like coherence, are construed as inductive fit and unification as constituted by (conditional) positive relevance relative to a belief function.

On a strategic level, the main difference between the papers referred to and the present proposal is that there fit and unification are somehow construed as relations between facts or phenomena and theories, whereas here they are construed as a relation between facts and inductive schemes. Talking of theories is certainly closer to scientific practice, but talking of inductive schemes is nearer to epistemological theory. Is there a substantial difference? This is unclear as long as the relation between theories and inductive schemes is not made clear. Without doubt there is a close relation, and it is incumbent on me to say how theories are implicitly contained in inductive schemes; I shall not attempt to do so here. Conversely, however, there is an urgent need

to say how inductive schemes are implicitly contained in scientific theories; I am convinced that mere reference to theories is not helpful for all the topics here addressed as long as theories are conceived as something modally inert, e.g. as sets of extensional sentences or extensional models.[47]

These remarks also suggest an answer to the question on which Salmon (1978) hangs his discussion. Salmon asks on p. 684:

> Suppose you had achieved the epistemic status of Laplace's demon . . . who knows all of nature's regularities, and the precise state of the universe . . . at some particular moment. . . . you would be able to predict any future occurrence, and you would be able to retrodict any past event. Given this sort of apparent omniscience, would your scientific knowledge be complete . . .? Laplace asked no more of his demon; should we place further demands upon ourselves?

In the sequel Salmon explains what Laplace's demon lacks. From the present point of view, omniscience — whether it is direct as presumably that of God or inferred from a complete set of axioms as that of Laplace's demon — is neither an ideal nor a counterfactual epistemological possibility for us. The reason is not that it is impossible on various scores to know so much. The reason is rather that we could not merely know everything; having an inductive scheme, proceeding inductively in the broad sense here always referred to is an essential and indispensible feature of our epistemic constitution which would not fade by approaching omniscience. Laplace's demon is indeed granted too little; it would not know what to believe, if it were to discover that it is wrong. We would know, even while approaching omniscience. If I am right, all the other things which the demon is held to be wanting in this discussion including those mentioned by Salmon (1978), p. 701, result from this central lack.[48]

Having thus shed some light on the epistemological locus of stable reasons, I can finally turn to the object of my paper: the relation between explanations and ultimately stable reasons. Though the definitions of $C_{A,B}$ (Def. 7) and of $S_{A,B}$ (Def. 8) look quite similar, this relation is not straightforward. The main difference is that direct causes are characterized by conditionalization on the whole past of the effect, whereas ultimately stable reasons are characterized by conditionalization on many and finally all other truths, whether past or future. This prevents a direct comparison. There is help, however: just restrict all the coherentistic considerations about the j-proposition B in the previous section to the past of B. This move brings easy success, indeed too easy, and therefore two disappointments. I shall first state in precisely what the move and its success consist, before explaining what the disappointments are and what might be done about them.

The move is simple: Restate PCo1 as saying that for any $j \in I^*$ and any contingent j-proposition B there is an $I^*_{<j}$-measurable reason for B relative to β^*. This is equivalent to a modified PCa1 saying that for any such j and B there is a direct cause of B in some world $\omega \in \Omega$ relative to β^*. Change PCa2 and PCo2 in the same manner; the former is again implied by the latter.[49] Modify finally Definition 8: Define A to be a *ω, j-past-stable* (*sufficient, necessary*) *reason for B within C relative to β* by additionally requiring C to be $I_{<j}$-measurable and by requiring A to be a (sufficient, necessary) reason for B relative to β conditional only on each $I_{<j}$-measurable $D \subseteq C$ with $\omega \in D$ and $A \cap D \neq \overline{A} \cap D$; define accordingly A's being an *ultimately ω, j-past-stable* (*sufficient, necessary*) *reason for B* and A's *range of being an ultimately j-past-stable* (*sufficient, necessary*) *reason for B*; and denote this range by $S_{j;A,B}({}^sS_{j;A,B}, {}^nS_{j;A,B})$. PCo3 may then be reformulated correspondingly.

After this modification the comparison is immediate: If A is an i-proposition and B a j-proposition with $i < j$ and if i is a binary variable,[50] then $A \cap B \cap C_{A,B} = S_{j;A,B}$ ($A \cap B \cap {}^sC_{A,B} = {}^sS_{j;A,B}$, $A \cap B \cap {}^nC_{A,B} = {}^nS_{j;A,B}$) and thus A causally explains B (as necessary, as possible) relative to β if and only if A is believed to be a j-past-stable (sufficient, necessary) reason for B relative to β. For proof one has only to look at the definitions and to observe first that ${}^{\omega}(I_{<j}\text{-}\{i\})$ is *the* smallest $I_{<j}$-proposition C with $\omega \in \Omega$ and $A \cap C \neq \emptyset \neq \overline{A} \cap C$, if i is binary, and secondly that being an ultimately ω,j-past-stable reason only requires being a reason conditional on this smallest proposition C.

Hence the justification of explanation I propose runs as follows: On the one hand, there is the explication of direct causation and consequently of causal explanation in section 3. On the other hand, there are the independent coherentistic considerations of section 4 which suggest that truth is tied to ultimately stable reasons, as stated in PCo3, and that believing in ultimately stable reasons is thus an indispensible ingredient of having a true world picture. And, as has turned out now, it is explanations and only explanations which provide these ingredients, at least if the relation of being a reason is considered only with respect to

pairs of logically simple propositions about single variables and if the coherentistic considerations are restricted to the past of the later proposition of such a pair.

Somehow, however, the last step appears too trivial. It falls short of the expectations I have created in two respects:

One disappointment is that in the short proof of the identity of $A \cap B \cap C_{A,B}$ and $S_{j;A;B}$ being an ultimately stable reason takes on an unexpectedly weak sense. According to my definition, being an ultimately ω(j-past)-stable reason boils down to being a reason conditional on, sloppily put, all the rest of the truth in (the j-past of) ω. But according to my motivation in the previous section, the idea rather was that an ultimately ω-stable reason is a reason after some finite information true in ω and stays a reason after all further information true in ω.

This deficiency can be cleared, however, because the cause specified in a causal explanation is in fact a reason which is stable within the cause's range and not only from some remote point onward. More precisely, the following assertion holds true: If A is an i-proposition and B a j-proposition with $i < j$, then A is a sufficient reason for B conditional on each non-empty, $I_{<j}$-$\{i\}$-measurable $D \subseteq {}^{s}C_{A,B}$ relative to β.[51] Hence, for each $\omega \in A \cap B \cap {}^{s}C_{A,B}$, A is a ω,j-past-stable sufficient reason for B not only ultimately, but within no less than ${}^{s}C_{A,B}$. The assertion with 'necessary' and '${}^{n}C_{A,B}$' replacing 'sufficient' and '${}^{s}C_{A,B}$' holds correspondingly.[52] However, the assertion fails to generally hold for reasons and direct causes simpliciter.[53]

Do explanations also provide unconditional reasons? Under mild assumptions yes, provided only sufficient or necessary reasons are considered. More precisely, the following assertion holds true: If A causally explains B as necessary relative to β and $\beta({}^{s}C_{A,B} | \overline{A}) \geqslant 0$, then A is a sufficient reason for B relative to β[54]; and if A causally explains B as possible relative to β and $\beta(B\text{-}{}^{n}C_{A,B} | \overline{A}) < 0$, then A is a necessary reason for B relative to β.[55] This assertion, or at least its 'sufficient'-part, very much resemble the thesis "that an adequate answer to an explanation-seeking why-question is always also a potential answer to the corresponding epistemic why-question"[56] and may thus be taken as a proof thereof. The additional premise of the 'sufficient'-part that $\beta({}^{s}C_{A,B} | \overline{A}) \geqslant 0$ will usually be satisfied, I think; and one might argue that it is just this premise which is violated in alleged counter-examples to that thesis.[57]

The other disappointment is the restriction of the coherentistic considerations about the j-proposition B to the past of B in the way specified above. This is disappointing because thus restricted these considerations lose much of their persuasiveness. I have great confidence in PCo1-3 as I have stated them in the previous section; but I do not know how to convincingly argue for PCo1-3 as modified in this section. The modified PCo1-3 (and in particular the principles of causality associated with them) still look desirable, but it is not clear why they should be necessary on coherentistic grounds. This is a gap in my argument.

Perhaps, however, this unsupported restriction of the coherentistic considerations is not really necessary. How is it possible that the i-proposition A is a direct cause of the j-proposition B in ω and thus an ultimately ω,j-past-stable reason for B, but not an ultimately ω-stable reason for B? The only possibility is that some true information about the future of B turns the positive relevance of A for B given the rest of the past of B into irrelevance or negative relevance. But there is something odd about this possibility. Consider a simple formal example:

Let ω, A, B, and C be such that A precedes B, B precedes C, and $\{\omega\} = A \cap B \cap C$. Now suppose on the one hand that A is a sufficient reason for B and thus also a sufficient direct cause of B in the small world ω, and on the other hand that A is a necessary reason for $\overline{B}$ given C and thus not an ultimately ω-stable reason for B. These assumptions imply: First, $\beta(C | A) < 0$; thus A and C cannot both believed to be true, and A is at best a weak reason for C. Secondly, C is a necessary reason for $\overline{B}$ given A and, because of symmetry, B a reason for $\overline{C}$ given A.[58] Hence, C very badly fits A and B under these assumptions.

Similar assumptions create similar oddities. This suggests a general conclusion which looks at least plausible: If for all $j \in I$ true j-propositions cohere with all past truths, then, for any $i \in I$, a true i-proposition coheres with all other truths, because it coheres with all past truths, as just stated, and also with true j-propositions for all $j > i$, since coherence is symmetric. In this way general coherence with the past seems to be equivalent with general coherence with past *and* future. If this is true, the above restriction of the coherentistic considerations would, after all, not really be a restriction. However, this is only a vague conjecture, neither precisely stated nor proved.

If the presented line of reasoning is correct, we ask 'why?', we search for explanations because this is one and, in a way, the only way of

finding coherent truth and, insofar as truth must be believable and coherent, the only way of finding truth. Why search for truth? Here I cannot think of any further theoretical justification; to some extent we seem to be intrinsically curious beings. Papers must end, justifications presumably, too. But the present one does not end here; there is beautiful further justification for the search for truth of a practical, decision-theoretic kind.[59]

Department of Philosophy,
University of Bielefeld,
Bielefeld, Germany.

NOTES

* I am very grateful to Dan Hunter for checking my English.

[1] Sketchy remarks about the utility of explanations may be found quite often. Much less often the question is explicitly addressed, e.g. by Salmon in (1978), where he propounds his own answer to the question, and in (1984), pp. 124ff. and 259ff., where he discusses also other answers.

[2] This is the declared strategy of, e.g., Friedman (1974) and Kitcher (1981).

[3] Indeed, this paper originated from an observation of this fit.

[4] Cf., e.g., Hume (1739), pp. 94ff., and (1777), pp. 47ff. The struggle is most conspicuous in the appendix of (1739), pp. 623ff.

[5] In (1777), p. 26, Hume writes: "All reasonings concerning matter of fact seem to be founded on the relation of *Cause and Effect*. By means of this relation alone we can go beyond the evidence of our memory and senses." In (1739), p. 107, he says equally clearly: ". . . we find by experience, that belief arises only from causation, and that we draw no inference from one object to another, except they be connected by this relation". However, the relevance of further principles of association, namely resemblance and contiguity, is not really clear. In (1739), pp. 107ff., he argues that these are only assisting, but not basic principles; in (1777), pp. 50ff., he does not discriminate in this way.

[6] This is very explicit in Hume (1739), pp. 170—172.

[7] This observation raises questions: Does Hume take one of the two notions as primary? Or is there a circularity in Hume's account? Which role has Hume's definition of causation taken as what he calls a philosophical relation, which refers to regularity instead of association? Does it offer a way out of the possible circularity? Cf., e.g., Mackie (1974), ch. 1, and Beauchamp, Rosenberg (1981), ch. 1, for thorough discussions of these questions.

[8] This is the lesson, for instance, of Goodman's new riddle of induction and Carnap's acknowledged failure to distinguish even a small class of inductive methods. It is challenged, however, by the puzzling alternative set up in the final section of Lewis (1980).

[9] Cf., e.g., Hunter (1991) in this volume.

[10] The further conclusion that plain belief is an illusion is unwarranted; it is drawn only in default of a more appropriate representation of epistemic states.

[11] See Shackle (1961), Rescher (1976), and Cohen (1977).

[12] He has summarized his work in Gärdenfors (1988).

[13] The only point of this technical label is that it be not confused with other notions. Perhaps the more suggestive term 'disbelief function' would be better, as Shenoy (1991) has proposed.

[14] My account in (1988) is slightly more general insofar as the range there consists of ordinal numbers.

[15] Setting $\kappa(\emptyset) = \infty$ is a reasonable convention. But ∞ should not be allowed as value of possible worlds and consistent propositions because no good rules of belief revision can be devised for it.

[16] The various problems which cast serious doubt on the idea that belief takes propositions as objects are pressing, but must here be disregarded.

[17] $\overline{A}$ of course denotes the complement of A relative to Ω.

[18] I thereby follow a proposal of Shenoy (1991).

[19] This definition which is much simpler than my original one has been pointed out to me by Bernard Walliser. Note that because of the law of negation at least one of the terms of the definiens is 0.

[20] The reason why the more vivid belief functions are introduced only as a derivative concept is that their formal behavior is less perspicuous.

[21] This holds because $\kappa(C|A \cup B) = \kappa((C \cap (A \cup B) - \kappa(A \cup B) = \min[\kappa(A \cap C, \kappa(B \cap C)| - \min[\kappa(A), \kappa(B)|$ and because $\min[y_1, y_2| - \min[x_1, x_2|$ is always between $y_1 - x_1$ and $y_2 - x_2$.

[22] In my (1988), p. 117, I have defined this process as A,n-conditionalization.

[23] For further details see my (1988) or my (1990b). There it is made clear why, given certain assumptions, revision schemes for plain belief have to take the form of NCFs; it is shown that NCFs behave very much like probability measures with respect to conditionalization and (conditional) independence; and the justification for more general forms of conditionalizations of NCFs closely parallels that for Jeffrey's generalized probabilistic conditionalization given by Teller (1976).

[24] A proposition A is contingent iff $\emptyset \neq A \neq \Omega$.

[25] This is Hume's most influential conclusion of the crucial Section XIV of Book I of his (1739). "The idea of necessity arises from some impression. . . . It must . . . be derived from some internal impression, or impression of reflexion", he writes on p. 165, and it is clear that necessity here includes causal necessity.

[26] If taken subjectively, Lewis' similarity relations are similar, but not equivalent to my NCFs; see my (1988), p. 127.

[27] My (1992) is an attempt to meet this obligation and thus to do justice to the realistic intuition within the present framework.

[28] In the third paragraph of this section I claimed to have a neutral usage of 'proposition'. This may seem objectionable because I assume propositions to be objects of belief *and* of causation and thus to play a double role which is arguably unsatisfiable. This is indeed a problem. But the problem arises within a realistic conception of causation and is thus part of and additional burden to the objectivization problem just put aside.

[29] It should be clear that a fact may have several causes. Thus I follow the common understanding which construes 'cause' as 'partial cause' and not as 'total cause'.

[30] This is precisely the idea and the conclusion of Cartwright (1979) concerning probabilistic causation.

[31] Even for direct causation it would not be reasonable to generally require that i immediately precedes j, because the frame I may be any wild collection of variables; in particular, I may contain many variables which are temporally, but not causally, between i and j. This requirement at best characterizes nice frames. It should be noticed that the assumption of discrete time is important for Definition 6; given continuous time, direct causes presumably do not exist or are simultaneous with their effects.

[32] Cf. my (1988), section 6.

[33] See the enlightening discussions of Bunzl (1979) and Lewis (1986), pp. 193–212.

[34] This, by the way, is also a point of difference between Gärdenfors' belief revision model and mine. Gärdenfors has plain conditional belief, but not more or less firm conditional belief and therefore nothing like additional reasons and causes. Cf. Gärdenfors (1988), ch. 3 and 9.

[35] However, I tend to join Lewis (1986), pp. 221ff., in thinking that there is no non-causal explanation of singular facts.

[36] This is essentially also what Lewis (1986), pp. 217ff., maintains, though he points out that knowing a cause of an event is not the only way of having information about the causation of that event. I neglect here the other ways.

[37] In section 4 of my (1990a) I have called $C_{A,B}$ the actually relevant circumstances of (the direct causal relation between) A and B in the widest sense.

[38] The need to consider explanations by additional or weak causes will not arise; thus I did not formally introduce them.

[39] It would be useful to extend my comparative remarks about causation to some remarks about how Definition 7 relates to other accounts of explanation; but this is beyond the scope of this paper. Just this much: my account seems to me to fit van Fraassen's theory in sect. 5.4 of his (1980) insofar as Definition 7 tries to say more about van Fraassen's relation of explanatory relevance for the case of direct causal explanation.

[40] It would not be reasonable to ask without restrictions on measurability whether there is an inductive reason for any contingent proposition B not of the form $\{\omega\}$ or Ω-$\{\omega\}$, because the answer turns out to be yes whenever Ω has more than four members.

[41] However, I don't know of any theory about the evolution of inductive schemes, i.e. about the change of belief functions, probability measures, or whatever *for changing frames*, except of conceiving it as generated by an underlying, more inclusive inductive scheme.

[42] Cf. also the quite similar remarks of Ellis (1979), pp. 9ff., about what he calls the ideal of completability.

[43] This is easily proved on the basis of two properties of conditional independence between sets of variables which are stated as assertion (7) in my (1990b) and proved as Theorems 11 and 13 in my (1988). A probabilistic counterpart of the present claim, or rather a considerable generalization thereof, is proved as Theorem 5 in my (1980).

[44] Which is the basic theme of Putnam's recent work; cf., e.g., the introduction and ch. 4 of Putnam (1983).

[45] Proof: Let the i-proposition A be a direct cause of B in α^* relative to β^*. If $D = {}^{\alpha^*}(I_{<j}\text{-}\{i\})$, this says that $\beta^*(B|A \cap D) > \beta^*(B|\overline{A} \cap D)$. Let $E_1 = A \cap D$, $E_2 = \overline{A} \cap D$, $E_3 = A \cap \overline{D}$, $E_4 = \overline{A} \cap \overline{D}$, and $E = \bigcup\{E_r : \beta^*(B|E_r) \geqslant \beta^*(B|E_1)\}$. The law of disjunctive conditions (after Definition 3) immediately implies that $\beta^*(B|E) > \beta^*(B|\overline{E})$, i.e. that E is a (unconditional) reason for B relative to β^*. The same reasoning applies if B has a direct effect in α^* instead of a direct cause.

[46] This is so because, as the proof in the previous note shows, there is an unconditional reason for B, whenever there is a conditional reason for B.

[47] On this score, then, the sentential and Sneed's and Stegmüller's structuralist view of theories seem equally insufficient. This insufficiency is also felt, for instance, by Kitcher (1981), when he associates explanatory stores of argument patterns with scientific theories. Cf. also Mühlhölzer (1989), ch. 6.

[48] Of course, the demon has other epistemological defects as well. For instance, it may be one of the two gods of Lewis (1979), pp. 502f., unable to localize itself. But this is obviously another kind of defect.

[49] The proofs given in the notes 43 and 45 also apply to these claims.

[50] This means that Ω_i has only two elements. This premise is technically required and I am not sure about the best way to get rid of it.

[51] Proof: For each $\omega \in {}^{s}C_{A,B}$ we have $\kappa(\overline{B}|A \cap {}^{\omega}(I_{<j}\text{-}\{i\})) > 0$ and $\kappa(\overline{B}|\overline{A} \cap {}^{\omega}(I_{<j}\text{-}\{j\})) = 0$. Trivially, each $I_{<j}$-$\{i\}$-measurable $D \subseteq {}^{s}C_{A,B}$ is equal to $\bigcup\{{}^{\omega}(I_{<j}\text{-}\{i\}) : \omega \in D\}$. Therefore, the law of disjunctive conditions (after Definition 3) implies the assertion that for each such D $\kappa(\overline{B}|A \cap D) > 0$ and $\kappa(\overline{B}|\overline{A} \cap D) = 0$.

[52] At this point it is particularly clear that my analysis of explanation is closely related to Hempel's requirement of maximal specificity (cf. Hempel (1965), pp. 397ff.) and to Skyrms' notion of resiliency (cf. Skyrms (1980) parts IA and IID).

[53] The failure of the analogous probabilistic assertion is related to Simpson's paradox. Cf. also my (1990a), pp. 128.

[54] Proof: Let ${}^{s}C_{A,B}$ be abbreviated by C. It was just shown that A is a sufficient reason for B conditional on C, i.e. that (a) $\kappa(\overline{B}|A \cap C) > 0$ and (b) $\kappa(\overline{B}|\overline{A} \cap C) = 0$. Since A causally explains B as necessary, A, B, and C are believed; this implies $\kappa(A) = 0$ and $\kappa(\overline{C}) > 0$; hence $\kappa(A \cap \overline{C}) > 0$ and (c) $\kappa(\overline{C}|A) = \kappa(A \cap \overline{C}) - \kappa(A) > 0$. And the additional premise says that (d) $\kappa(C|\overline{A}) = 0$. Now, (a) implies $\kappa(\overline{B} \cap C|A) > 0$, and (c) implies $\kappa(\overline{B} \cap \overline{C}\ A) > 0$; therefore $\kappa(\overline{B}\ A) > 0$. Moreover, (b) and (d) imply $\kappa(\overline{B} \cap C|\overline{A}) = 0$ and thus $\kappa(\overline{B}|\overline{A}) = 0$.

[55] Proof: Let ${}^{n}C_{A,B}$ be abbreviated by C. It was just shown that A is a necessary reason for B conditional on C, i.e. that (a) $\kappa(B|A \cap C) = 0$ and $\kappa(B|\overline{A} \cap C) > 0$, hence (b) $\kappa(B \cap C|\overline{A}) > 0$. Since A causally explains B as possible, A, B, and C are believed; thus $\kappa(A \cap C) = 0$ which implies (c) $\kappa(C|A) = 0$. And the additional premise says that (d) $\kappa(B \cap \overline{C}|\overline{A}) > 0$. Now, (a) and (c) imply $\kappa(B \cap C|A) = 0$ and thus $\kappa(B|A) = 0$. Moreover, (b) and (d) imply $\kappa(B|\overline{A}) > 0$.

[56] Hempel (1965), pp. 368. This is the part the thesis of the structural identity of explanation and prediction which Hempel (1965), pp. 364ff., endorses.

[57] I have in mind Michael Scriven's examples of the jealous murderer and the collapsing bridge discussed in Hempel (1965), pp. 370ff.

[58] For proof note that (a) $\kappa(\overline{B}|A) = \min[\kappa(\overline{B}|A \cap C) + \kappa(C|A), \kappa(\overline{B}|A \cap \overline{C}) + \kappa(\overline{C}|A)]$. We have assumed (b) $\kappa(\overline{B}|A) > 0$ and (c) $\kappa(\overline{B}|A \cap C) = 0$. All three

immediately imply the first claim $\kappa(C|A) > 0$. According to the law of negation (after Definition 1), the latter entails $\kappa(\overline{C}|A) = 0$; this and (a) in turn entail $\kappa(\overline{B}|A \cap \overline{C}) > 0$; and this and (c) say that C is a necessary reason for $\overline{B}$ given A.

[59] I refer to the observation in Savage (1954), sect. 7.3, that the expected utility of free information is always non-negative, and to the strong generalizations offered by Skyrms (1990), ch. 4. A different generalization to free memory may be found in Spohn (1978), sect. 4.4.

REFERENCES

Beauchamp, Tom, Rosenberg, Alexander: 1981, *Hume and the Problem of Causation*, Oxford University Press, Oxford.

Bunzl, Martin: 1979, 'Causal Overdetermination', *Journal of Philosophy* **76**, 134—150.

Cartwright, Nancy: 1979, 'Causal Laws and Effective Strategies', *Noûs* **13**, 419—437.

Cohen, Jonathan L.: 1977, *The Probable and the Provable*, Clarendon Press, Oxford.

Ellis, Brian: 1979, *Rational Belief Systems*, Blackwell, Oxford.

Field, Hartry: 1972, 'Tarski's Theory of Truth', *Journal of Philosophy* **69**, 347—375.

van Fraassen, Bas C. 1980, *The Scientific Image*, Oxford University Press, Oxford.

Friedman, Michael: 1974, 'Explanation and Scientific Understanding', *Journal of Philosophy* **71**, 5—19, German translation in Schurz (1988), pp. 171—191.

Gärdenfors, Peter: 1988, *Knowledge in Flux*, MIT Press, Cambridge, Mass.

Hempel, Carl Gustav: 1965, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, Free Press, New York.

Hume, David: 1739, *A Treatise Concerning Human Nature*. Page numbers refer to the 2nd edition by P. H. Nidditch of the edition of L. A. Selby-Bigge, Oxford University Press 1978, Oxford.

Hume, David: 1777, *An Enquiry Concerning Human Understanding*. Page numbers refer to the 3rd edition by P. H. Nidditch of the edition of L. A. Selby-Bigge of David Hume's Enquiries, Oxford University Press 1975, Oxford.

Hunter, Daniel: 1991, 'Maximum Entropy Updating and Conditionalization', this volume, pp. 45—57.

Kitcher, Philip: 1981, 'Explanatory Unification', *Philosophy of Science* **48**, 507—531, German translation in Schurz (1988), pp. 193—229.

Lambert, Karel: 1988, 'Prolegomenon zu einer Theorie des wissenschaftlichen Verstehens', in Schurz (1988), pp. 299—319.

Lambert, Karel: 1991, 'On Whether an Answer to a Why-Question is an Explanation if and only if it Yields Scientific Understanding', in: G. G. Brittan jr. (ed.), *Causality, Method, and Modality*, Kluwer, Dordrecht, pp. 125—142.

Lewis, David: 1973, 'Causation', *Journal of Philosophy* **70**, 556—567; also in Lewis (1986), pp. 159—172.

Lewis, David: 1979, 'Attitudes de dicto and de se', *Philosophical Review* **88**, 513—543.

Lewis, David: 1980, 'A Subjectivist's Guide to Objective Chance', in R. C. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, vol. II, University of California Press, Los Angeles, pp. 263—293; also in Lewis (1986), pp. 83—113.

Lewis, David: 1986, *Philosophical Papers*, Vol. II, Oxford University Press, Oxford.

Mackie, John L.: 1974, *The Cement of Universe*, Oxford University Press, Oxford.

Mühlhölzer, Felix: 1989, *Objektivität und Erkenntnisfortschritt. Eine Antwort auf Thomas S. Kuhn*, Habilitationsschrift, Munich.

Putnam, Hilary: 1983, *Realism and Reason, Philosophical Papers*, Vol. 3, Cambridge University Press, Cambridge.

Rescher, Nicholas: 1973, *The Coherence Theory of Truth*, Oxford University Press, Oxford.

Rescher, Nicholas: 1976, *Plausible Reasoning*, Van Gorcum, Assen.

Rescher, Nicholas: 1985, 'Truth as Ideal Coherence', *Review of Metaphysics* **38**, 795—806.

Salmon, Wesley C.: 1970, 'Statistical Explanation', in: R. G. Colodny (ed.), *Nature and Function of Scientific Theories*, University of Pittsburgh Press, Pittsburgh, pp. 173—231.

Salmon, Wesley C.: 1978, 'Why Ask, 'Why?'? An Inquiry Concerning Scientific Explanation', *Proceedings and Addresses of the American Philosophical Association* **51**, 683—705.

Salmon, Wesley C.: 1984, *Scientific Explanation and the Causal Structure of the World*, Princeton University Press, Princeton.

Savage, Leonard J.: 1954, *The Foundations of Statistics*, Wiley, New York.

Schurz, Gerhard (ed.): 1988, *Erklären und Verstehen in der Wissenschaft*, Oldenburg, Munich.

Shackle, G. L. S.: 1961, *Decision, Order, And Time in Human Affairs*, Cambridge University Press, Cambridge.

Shenoy, Prakash: 1991, 'On Spohn's Rule for Revision of Beliefs', *International Journal of Approximate Reasoning* **5**, 149—181.

Skyrms, Brian: 1980, *Causal Necessity*, Yale University Press, New Haven.

Skyrms, Brian: 1990, *The Dynamics of Rational Deliberation*, Harvard University Press, Cambridge, Mass.

Spohn, Wolfgang: 1978, *Grundlagen der Entscheidungstheorie*, Scriptor, Kronberg/Ts.

Spohn, Wolfgang: 1980, 'Stochastic Independence, Causal Independence, and Shieldability', *Journal of Philosophical Logic* **9**, 73—99.

Spohn, Wolfgang: 1983, 'Deterministic and Probabilistic Reasons and Causes', in C. G. Hempel, H. Putnam, W. K. Essler (eds.), *Methodology, Epistemology, and Philosophy of Science. Essays in Honour of Wolfgang Stegmüller on the Occasion of his 60th Birthday*, Reidel, Dordrecht, pp. 371—396.

Spohn, Wolfgang: 1988, 'Ordinal Conditional Functions. A Dynamic Theory of Epistemic States', in: W. L. Harper, B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics*, Vol. II, Kluwer, Dordrecht, pp. 105—134.

Spohn, Wolfgang: 1990a, 'Direct and Indirect Causes', *Topoi* **9**, 125—145.

Spohn, Wolfgang: 1990b, 'A General Non-Probabilistic Theory of Inductive Reasoning', in: R. D. Shachter, T. S. Levitt, L. N. Kanal, J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence* **4**, Elsevier, Amsterdam, pp. 149—158.

Spohn, Wolfgang: 1992, 'Causal Laws are Objectifications of Inductive Schemes', in: J. Dubucs (ed.), *Theory of Probability*, Kluwer, Dordrecht, to appear.

Teller, Paul: 1976, 'Conditionalization, Observation, and Change of Preference', in: W.

L. Harper, C. A. Hooker (eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Vol. I, Reidel, Dordrecht, pp. 205–259.
Tooley, Michael: 1987, *Causation*, Oxford University Press, Oxford.

THE SYSTEMS OF PLATO AND ARISTOTLE COMPARED AS TO THEIR CONTRIBUTIONS TO PHYSICS

The science of kinematics did not arise before Christian Huyghens was able to analyse circular motion, though this analysis was implicit in Descartes' reflections and, from the time of Eudoxus and especially Ptolemy on, ingenious techniques concerning the measurement of superimposed or deformed circular motions had been developped in celestial kinematics. As to dynamics, if the law of falling bodies is due to Galileo and the principle of inertia is conceived in its generality by Descartes, the principles of motion find their first systematic expression in Newton's *Principia*. Newton's principles and their development during the 19th century gave to the relation between kinematics and dynamics a form which has been questioned by quantum mechanics. I aim to seek in the philosophical analysis of the concept of motion by Plato and Aristotle some explanations of the difficulties discovered in the relations between kinematics and dynamics and more generally in the history of physics.

I. A COMPARISON BETWEEN PLATO'S AND ARISTOTLE'S CONCEPTS OF MOTION

It is difficult to definitely assess Plato's proper concept of motion because of the impossibility of making a clear distinction between the statements advocated by the protagonists of the Dialogues and Plato's own theory. Moreover the genuine doctrine is often revealed and concealed in the guise of mythical expression. Nevertheless the following comparison may safely be drawn between the two greatest philosophers of Antiquity:

	PLATO	ARISTOTLE
I	1. The first origin of any motion or change is selfmotion, called also soul or life.	1. There is no self motion. Every motion is *ab alio*.

Wolfgang Spohn et al. (eds.), Existence and Explanation, 197–206.