

WHY FRIEDMAN'S NON-MONOTONIC REASONING DEFIES HEMPEL'S COVERING LAW MODEL

ABSTRACT. In this paper we will show that Hempel's covering law model can't deal very well with explanations that are based on incomplete knowledge. In particular the symmetry thesis, which is an important aspect of the covering law model, turns out to be problematic for these explanations. We will discuss an example of an electric circuit, which clearly indicates that the symmetry of explanation and prediction does not always hold. It will be argued that an alternative logic for causal explanation is needed. And we will investigate to what extent non-monotonic epistemic logic can provide such an alternative logical framework. Finally we will show that our non-monotonic logical analysis of explanation is not only suitable for simple cases such as the electric circuit, but that it also sheds new light on more controversial causal explanations such as Milton Friedman's explanation of the business cycle.

1. INTRODUCTION

There are different modes of explanation of which Hempel's *covering law model* has received most attention. For a long period of time philosophers of science thought that the covering law model is essentially correct. This attitude might be partly explained by the fact that its formalization has proved to be possible in a logical satisfactory way. In recent years Hempel's model of explanation has been severely criticized, and interest seems to shift away from logical analyses of explanation (see Salmon 1989). In this paper we extend the list of criticisms by claiming that the covering law model is not very well suited to deal with situations in which there is incomplete knowledge (information). The covering law model can only be applied to cases where everything is known: the relevant empirical law, initial conditions, absence of potential distorting factors, etc. We claim that in science these conditions are seldom met. This paper focuses attention upon situations in which the relevant covering law is not known. We will show that in cases of incomplete knowledge intuitive appealing explanations are still feasible. In addition we offer a formalization of this type of explanation in so called non-monotonic logic, which is supplementary to the deductive scheme of Hempel's covering law model. Non-monotonic logic has been recently developed in artificial

intelligence research to formalize reasoning with incomplete information. We will show that in particular the non-monotonic epistemic logic (NMEL) that was introduced by Y. Shoham is very suitable to formalize explanations based on incomplete knowledge. So, although we agree with much of the criticism on Hempel's model, we still think a logical analysis of explanation is feasible.

A second purpose of the paper is to give an alternative account of the method Milton Friedman actually uses in his studies on monetary history. We will see that Friedman's work is permeated with passages which indicate that he is fully aware of the fact that in economics a lot is unknown. We will show that Friedman's explanations can be seen as paradigmatic examples of incomplete knowledge explanations. His explanatory arguments are formally reconstructed in NMEL. By offering such a formal reconstruction we do *not* claim to give a *justification* of Friedman's method. Our main objective is to show how he makes his case. (Of course, we would not have undertaken the effort to reconstruct his arguments, if we were of the opinion that there are no useful elements in it).

At this point an aside is in order. The methodology of economics has seen a vast literature on the exegesis of Friedman's (1953) essay in which he proposes an instrumentalist '*as-if*' methodology. The main thesis of this essay is that a theory should be tested by its predictions and not by the realism of its assumptions. We do not contribute to the exegesis of the essay. On the contrary, one of our purposes is to show that the '*as-if*' methodology debate, or the F-twist, has directed the attention to issues that have hardly contributed to an understanding of the method Friedman actually applies in his economic work.¹ Traditionally, Friedman's research as an economist has been interpreted as an implementation of his own '*as-if*' methodology (see, e.g., Blaug 1976). We claim that Friedman's main contributions to economics can be better understood as a case of incomplete knowledge explanations. This does not mean that we deny that aspects of the '*as-if*' methodology can be pointed out in Friedman's research, but instead we simply believe that those aspects are not as important as other authors seem to suggest.

The paper is organized as follows. In Section 2 we present an example of an electric circuit in which a large part of the circuit is hidden (unknown) to the observer. The example serves a dual purpose. It elucidates the concept of incomplete knowledge, and at the same time, some more substantial points about explanation and prediction based

on incomplete knowledge can be learned from it. Among other things it is shown that there is an asymmetry between explanation and prediction which cannot be accounted for in the covering law model. In Section 3 we demonstrate that there are many striking similarities between the electric circuit example and Friedman's work. For this purpose we analyze Friedman and Schwartz's *Monetary History* (1963, and Friedman 1969). In order to taste something of the flavor with which Friedman presents his arguments we give several quotations from his original work. Section 4 presents Shoham's non-monotonic logic NMEL and its application to the electric circuit example. It turns out that within NMEL a natural account of the asymmetry between explanation and prediction can be given. In Section 5 we apply NMEL to Friedman's studies on monetary history, and it is shown that our NMEL analysis contributes to a better understanding of Friedman's arguments. Finally, in Section 6 the results of this paper will be summed up, and we give some suggestions for further research.

2. EXPLANATIONS WITHOUT COVERING LAWS

In Tan (1988) an example of an explanation with incomplete information is discussed which is not covered by Hempel's *Covering Law Model* of causal explanations.² In this section we will study this 'counter'-example more closely. According to Hempel an explanation is a deductive inference which can be roughly presented as follows:

$$\begin{array}{l}
 \text{(I)} \quad (1) P(a) \\
 \quad \quad (2) \forall x(P(x) \rightarrow O(x)) \\
 \hline
 Q(a)
 \end{array}$$

$Q(a)$ is an observed event, which is caused by the occurrence of $P(a)$. The second premise $\forall x(P(x) \rightarrow Q(x))$ is a covering law, which constitutes the core of the deductive explanation. Covering laws are usually derived from a more general theory. An important aspect of Hempel's deductive account of explanations is the symmetry between explanation and prediction. The *symmetry thesis* says that if we can deductively explain with hindsight (*ex post*) the occurrence of $Q(a)$ from premises 1 and 2, then we could as well have predicted in advance (*ex ante*) that $Q(a)$ will occur if $P(a)$ occurs. This symmetry can be illustrated with the following example:

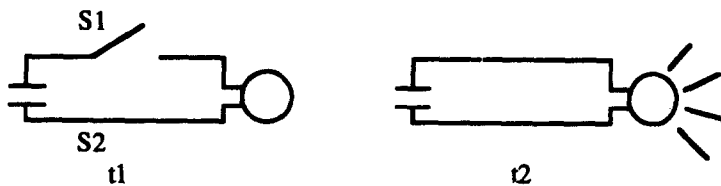


Fig. 1.

This figure represents an electric circuit at two consecutive stages, t_1 and t_2 . At timepoint t_1 the switch S_1 is off and the lamp L is off too. At the next moment t_2 the switch S_1 is turned on, with the consequence that the lamp goes on. This causal relation can be expressed by the following law:

$$(W) \quad \forall t [((t, S_1 = \text{on}) \wedge (t, S_2 = \text{on})) \rightarrow (t, L = \text{on})].$$

Substituting this law for the second premise in Hempel's scheme, it is clear that the burning of the lamp is deductively explainable. Moreover it is also obvious that the symmetry thesis holds in this case. When we know in advance that at t_2 switch S_1 will be turned on and that S_2 will still be on, we could use W to predict that at t_2 the lamp will be on. Though obvious this symmetry thesis seems to be, there are cases in which the symmetry breaks down. Consider the next counter example:

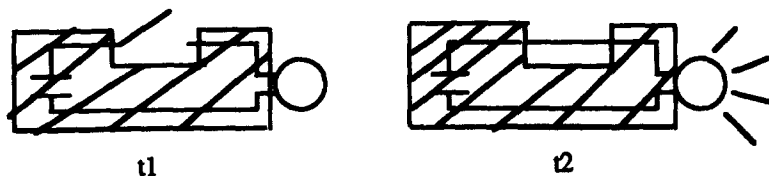


Fig. 2

The situation in Figure 2 is similar to Figure 1, except that now the circuit is contained in a black box and only S_1 and the lamp are observable from the outside. Suppose person P doesn't know what is in the box. At t_2 he turns S_1 on, and observes that the lamp goes on. When asked for an explanation, nobody will be surprised if P answers that the pressing of S_1 caused the burning of the lamp. This seems to be a perfectly normal causal explanation, but from a Hempelian point of

view something strange is going on here. First of all there is no symmetry. P can explain with hindsight why the lamp went on at t_2 , but he could not have predicted this event, as he does not know what is in the box. The box might have been empty, i.e., no battery and no wires at all. To know the law W, one has to know the interior of the box. And without knowing W, one cannot predict the burning of the lamp. Though P's ignorance accounts for the fact that he could not predict this event, it leaves P's explanation a mystery. According to Hempel's scheme, knowledge of covering laws is essential for explanations. However, the example above clearly shows not only that the symmetry thesis does not always hold, but also that causal explanations are not necessarily based on covering laws.

This phenomenon of asymmetry has been noticed before by other authors. A nice example of this asymmetry is discussed in Cartwright (1983, p. 51, 52). She knew that on the one hand camellias flourish when planted in rich soil, while on the other hand camellias die when planted in warm soil. Hence, when she actually planted her camellias in soil that was both rich and warm, she couldn't predict beforehand whether her camellias would flourish or die. After a while she observed that most of the camellias died. It was only after she made this observation that she was able to conclude that the camellias died, because they were planted in warm soil. She comments that this explanation is not an explanation from any true covering law. There is no law that says that camellias just like hers, planted in soil which is both hot and rich, die. Hence, she claims that "... our ability to give this explanation precedes our knowledge of that law". Although the role of incomplete knowledge in Cartwright's example is not so manifest as in our example, her example does illustrate very clearly that an *ex post* explanation does not always coincide with an *ex ante* prediction. Finally, an example which illustrates not only the asymmetry but also the role of incomplete knowledge is the crash of the stock market of October 1987. Hardly anybody had predicted that this crash would occur at the nineteenth of October. But once the crash had happened, economists were eager to explain that it was bound to happen.

Our observation that we can have explanations without covering laws should not be misunderstood. We do not claim that knowledge about electricity is irrelevant for P's explanation. Some general background knowledge does play a role in P's explanation, and a very important role indeed! If P would have answered that simply touching the switch

caused the lamp to burn, his explanation wouldn't be taken seriously, because it indicates that he doesn't know anything about electricity. Or if P were very imaginative he might even have reversed the causal direction; e.g., he might have answered that some Powerful God, by turning on the light, forced him to press the switch. Obviously these explanations are not acceptable, because they are not in accordance with our general knowledge about electricity. However, this kind of knowledge plays a totally different role than the empirical law W in Figure 1. This general background knowledge is used to narrow down the set of causal laws that could account for the burning of the light; i.e., it acts like a set of boundary conditions for the adequacy of explanations. Before P observed that pressing S_1 was followed by the burning of the lamp, it was still possible that there were no wires in the box and henceforth that there was no causal link whatsoever between S_1 and the lamp. And it was only after this observation that he could exclude this possibility. As the two events of turning on the switch and the lamp occurred at virtually the very same moment, P could in principle still have been in doubt about the direction of causality. But the possibility that pressing the switch was caused by the burning of the lamp was immediately ruled out by his background knowledge. Thus P arrived at the 'only' sensible conclusion that was left; i.e., turning on the switch caused the lamp to burn. This suggests the following type of argument:

$$\begin{array}{ll}
 \text{(II)} & (1) P(a) \\
 & (2) Q(a) \\
 & \text{-----} \\
 & \forall x(P(x) \rightarrow Q(x))
 \end{array}$$

Prima facie this scheme looks rather curious. It is a non sequitur in classical logic, as is indicated by the dotted line. However, this problem will be dealt with in Section 4. There we will see that the non-monotonic logic NMEL validates inferences like the scheme above. Applying scheme II to P's explanation of the fact that the lamp went on, it is obvious that P is not in the position to conclude W. Because W contains the necessary condition that switch S_2 has to be on, and P is not even aware of the existence of this second switch. However, P could come up with a close approximation of W, i.e., something like

$$(W') \quad \forall t[((t, S_1 = \text{on}) \wedge (t, \neg D)) \rightarrow (t, L = \text{on})].$$

The expression $\neg D$, which stands for 'there are no distorting factors in the system under consideration', is a kind of *unspecified ceteris paribus clause*. P assumes that there is some sort of electric circuit hidden in the black box which is not influenced by distorting factors in the sense that the current is somewhere interrupted in the circuit, or that the power supply is not functioning properly.³ When this *ceteris paribus clause* holds, then pressing S_1 causes the lamp to burn. Thus we arrive at the following instantiation of scheme II:

$$\begin{array}{l}
 \text{(II')} \quad (1) (2, S_1 = \text{on}) \\
 \quad \quad (2) (2, L = \text{on}) \\
 \hline
 \forall t[(t, S_1 = \text{on}) \wedge (t, \neg D) \rightarrow (t, L = \text{on})]
 \end{array}$$

From the observation that at moment t_2 the lamp went on when he turned on the switch, P concluded that pressing the switch caused the burning of the lamp.

The moral of our analysis of the electric circuit example can be summed up as follows. The real problem of a Hempelian style deductive explanation is not performing the deduction itself; i.e., explanation is not just a logical exercise, but the real problem is to *find* the specific law that does the job. The basic problem is how to apply a general theory to a specific situation. Hempel also paid attention to this problem. According to Hempel the finding of a specific law that applies to a specific situation is also a deductive argument. Let T be the general theory about electricity; i.e., the set of all basic laws of electricity and let F be the factual description of the electric circuit; i.e., F mentions that the circuit consists of four components which are connected in a certain way by wires, etc. One has found the specific law W, if one can derive W from T and F, i.e., $T \cup F \vdash W$. However, this simple deductive account of finding a law does not work for the explanation of P. As P has only incomplete information about the electric circuit, he is unable to derive the specific law W, even if he would know T by heart. Scheme II' provides the best guess that P can come up with. Hence, scheme II is supplementary to Hempel's deductive account of explanation in the sense that it applies to cases with incomplete information which are not very well dealt with by Hempel's scheme.

We do not consider scheme II as some sort of substitute for an inductive logic to discover new empirical laws. We pointed out that scheme II only makes sense against the background of a well-established

general background theory T. Without the boundary conditions provided by T, arguments like II could produce unintended conclusions such as the 'Powerful God' explanation. We have assumed the existence of such a background theory. The question how to discover such general theories is beyond the scope of this article. Scheme II is not meant as a scheme for abductive reasoning either. In abductive reasoning it is argued that one 'infers' from the two premises $\forall x(P(x) \rightarrow Q(x))$ and $Q(a)$, that $P(a)$ is the most probable candidate to have caused $Q(a)$. However, in scheme II it is the causal law that is inferred and not the fact $P(a)$.

In the next section we will discuss one of Friedman's most famous studies in economics. We will see that Friedman's line of reasoning in his studies on monetary history is a clear example of an explanatory argument which is in accordance with scheme II.

3. FRIEDMAN ON MONETARY HISTORY

The example discussed in the previous section is something of a toy problem. In this section we show that a piece of actual scientific research exhibits the most substantial points made in the previous section. We will discuss Friedman and Schwartz's interpretation of the economic history of the U.S. from 1867 to 1960 as it is presented in Friedman (1969, chapter 10).

Friedman and Schwartz's *Monetary History* is devoted to an explanation of the fluctuations in some major economic quantities such as money and income.⁴ The chapter we discuss consists of three parts each of which presents a separate part of Friedman's explanation. (For a more detailed philosophical study of the literature concerning the *Monetary History* see Hirsch and de Marchi (1986)). In the first part Friedman presents some data about the cyclical behavior of the money stock (see Figure 3). At first sight there seems to be no relation between money and income (the latter as an indicator of economic activity, i.e., the business cycle). However, after presenting the data in terms of the rate of change Friedman observes a striking similarity in the rate of change of income and the stock of money. From this observation Friedman draws the conclusion that there "unquestionably is a close relation between the variability of the stock of money and the variability of income" (p. 205-6). However this 'close relation' is not a perfect correlation. According to Friedman the deviations in the relation be-

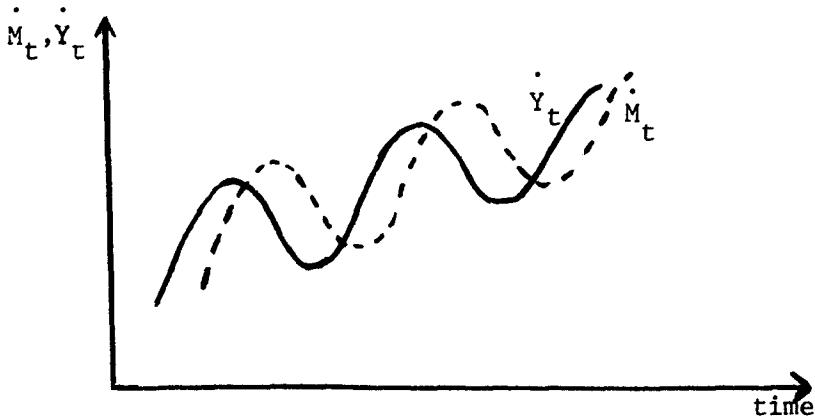


Fig. 3.

tween the rate of change of money and income can be explained away by pointing out distorting factors such as the “changing statistical character of the net national product estimates” (p. 203) and the “statistical errors in our money series” (p. 199). Furthermore, “standard deviations based on only four observations (three degrees of freedom) are subject to a good deal of sampling variation” (p. 203) and also, nominal income is “an inadequate index of economic change” (p. 199). As all the available evidence can be interpreted in such a way that deviations from the perfect correlation are accounted for by the presence of distorting factors, Friedman concludes that this correlation holds.

In the second part Friedman investigates what the direction of causality (Friedman uses the term influence) is: Are monetary changes caused by changes in economic activity, or are changes in economic activity caused by monetary changes? Friedman attacks this question by studying the historical circumstances underlying the changes in the stock of money. He shows that major changes in the stock of money can be attributed to specific autonomous historical circumstances. (According to Friedman, these circumstances are autonomous in the sense that they are not themselves causally influenced by the business cycle.) For this purpose Friedman uses the known fact that “changes in the stock of money can, arithmetically, be attributed to changes in three proximate determinants” (p. 209). (Here, arithmetically means “more or less by definition”.) He subsequently shows that each of the major

changes in the money stock is attributable to a change in one specific determinant, implicitly assuming that the other two determinants have not changed. He also shows that turning points in the rate of change of the money stock show a lead over turning points in the rate of change in income. From these two sources of evidence Friedman concludes that it is likely that the direction of causality runs from money to income. Thus, he arrives at a monetary account of the business cycle; i.e., an increase (decrease) in income is *caused* by an increase (decrease) in the stock of money. However, the evidence for a monetary account of the business cycle is not so strong for minor changes in the stock of money as it is for major changes. But this does not restrain Friedman from proposing his monetary explanation: "clearly, the view that monetary change is important does not preclude the existence of other factors that affect the course of business" (p. 222) and somewhat rhetorically he adds "[i]s not a common explanation for both (major and minor fluctuations) more appealing than separate explanations?" (p. 223). Thus, for minor changes Friedman does not commit himself to his monetary explanation. He makes as many reservations as possible without questioning the direction of causality, or in our terminology, he introduces an unspecified *ceteris paribus* clause.

In the account of the electric circuit we have given in the previous section a similar passage as the second part of Friedman's explanation seems to be largely absent at first sight. This is due to the fact that the 'Powerful God' explanation of the previous section is immediately ruled out, because it is not in accordance with the common sense belief that pressing the switch is an autonomous act with respect to the electric circuit. In Friedman's case the direction of causality between money and income is not so obvious and, therefore, it deserves a separate analysis. He shows that changes in the rate of growth of the money stock can be attributed to factors that may be regarded as autonomous with respect to the relation between money and income.

In the previous section we have remarked that P's explanation is convincing, because it is embedded in a system of background knowledge. The common sense component of this background knowledge consists of the following two elements: (i) we all know that switches and lamps stand in some relation to each other and (ii) our conjectures about the relation between switches and lamps have shown to be rather reliable in a lot of different situations. Precisely these two elements are also mentioned by Friedman: (i) "we know that money is a pervasive

element in the economy" (p. 214) and (ii) "the relation between money and business has remained largely unchanged over a period that has seen such substantial changes in the arrangements determining the quantity of money" (p. 215).

In the third part Friedman discusses the theoretical transmission mechanism that relates monetary and economic changes. A mechanism that is consistent with fundamental economic theory is needed, because "however consistent may be the relation between monetary and economic change, and however strong the evidence for the autonomy of monetary changes, we shall not be persuaded, unless we can specify in some detail the mechanism that connects the one with the other" (p. 229). That is, the proposed relation should also be in consonance with the economic theory contained in the background knowledge. For our purpose it is not needed to enter into the details of the proposed mechanism. It is sufficient to say that if the theoretical mechanism is the "true" mechanism relating monetary and economic changes then the covariance is no longer a surprising fact. The mechanism Friedman proposes is only tentative and of a *ceteris paribus* character: "let us neglect for the time being the effect of returns on other assets as well as still other possible variables" (p. 224).

The last part of Friedman's argument in the *Monetary History* is strongly related to the fact that P's explanation is consistent with (can be derived from) universal electricity laws.⁵ Of course as both systems are incompletely known one can only expect to derive conjectural knowledge, i.e., the proposed theoretical transmission mechanism may not be the actual one.

In the previous section we mentioned that P concludes the law W' from the observations that the lamp went on when the switch S_1 was turned on. In the same spirit Friedman infers from his observations concerning the cyclical pattern of the economy that the cyclical behavior of the money stock causes the cyclical behavior in major economic variables, notably income. This causal law relating money and income is the conclusion rather than a premise of Friedman's argument. A second point is that Friedman does not attempt to predict the future course of the business cycle on the basis of the covariance between the rate of change of the money stock and the rate of change of income. Friedman tries to interpret the evidence on money and income: he is engaged in explaining *ex post* a part of the economic history of the U.S. (i.e., 'rationalizing known evidence' – see Hirsch and de Marchi

(1986, p. 2) and Hammond (1988) and not in *ex ante* prediction. It is very questionable whether Friedman could have predicted this causal relation between money and income before he had the data at his disposal. Remember that in the first part of his study Friedman had to explain away many deviations in order to convince his readers that there was a correlation between the rate of change of the money stock and the rate of change of income. Apparently even this correlation was something of a surprise. This too is similar to the electric circuit example where person P cannot predict what will happen when he turns the switch S_1 on, but he is indeed able to explain *ex post* why the lamp went on when he turned on the switch.

4. NON-MONOTONIC EPISTEMIC LOGIC (NMEL)

Non-monotonic epistemic logic (NMEL) was introduced in Shoham (1988a, 1988b). It was developed to formalize reasoning with incomplete information.⁶ Before we can show how P's explanation of the electric circuit can be analyzed in NMEL, we first have to explain some technical details of NMEL. In this section we will only give a brief introduction of NMEL; for precise definitions and some relevant theorems the reader is referred to the appendix. In NMEL we can distinguish two aspects: an epistemic aspect and a non-monotonic aspect.

Epistemic aspect of NMEL

NMEL is a first-order predicate logic that contains two epistemic operators:

- $\Box\varphi$: P knows that φ is the case.
- $\Diamond\varphi$: P can assume that φ is the case, unless he knows that φ is not the case.

For example the formula $\Box(2, S_1 = \text{on})$ expresses that person P knows that the switch S_1 is on at timepoint t_2 . The \Diamond -operator is definable in terms of the \Box -operator:⁷

$$\Diamond\varphi =_{df} \neg\Box\neg\varphi.$$

Non-monotonic aspect of NMEL

From a semantic point of view the crucial difference between NMEL and a classical (monotonic) logic is the *minimality condition* on models. In a monotonic logic an argument is valid if the conclusion is true in all models in which the premises are true:

$A \vDash B$, i.e., A entails B , if B is true in all models of A .

In NMEL, however, only those models are taken into account that are minimal in some sense:

$A \vDash_C B$, if B is true in all *minimal* models of A .

(More formally: $A \vDash_C B =_{df}$ for all models M : if $M \vDash_C A$, then $M \vDash_C B$.)

Here, $M \vDash_C A$ means that M is a minimal model of A . Different definitions of the notion of minimality yield different non-monotonic logics. In NMEL the notion of minimality is defined as follows. M is a *minimal knowledge model* of the set of premises A , written $M \vDash_C A$, if

- (1) M satisfies A , i.e., $M \vDash A$.
- (2) for every alternative model M' that satisfies less formulas of the form $\Box\varphi$ than M ; it holds that M' does not satisfy A .

Condition 2 implies that a minimal model M of A satisfies as *few* formulas of the form $\Box\varphi$ as possible. Consider the example that $A = \{\Box(1, p), \Box(1, q)\}$. The only formulas of the form $\Box\varphi$ that are satisfied by a minimal model M of A are $\Box(1, p)$, $\Box(1, q)$ and the *classical* S5 deductive consequences of these formulas such as for example the conjunction $\Box(1, p) \wedge \Box(1, q)$.⁸ An example of a non-minimal model M' of A is a model that satisfies not only $\Box(1, p)$ and $\Box(1, q)$, but also $\Box(1, r)$, where $\Box(1, r)$ is not a classical S5 deductive consequence of $\Box(1, p)$ and $\Box(1, q)$. M' is not minimal, because $\Box(1, r)$ is true in M' , and there is an alternative model, namely M , that satisfies less formulas than M' and satisfies A as well.

The semantics of NMEL is a special case of a Kripke style possible world semantics for modal logic.⁹ A NMEL-model is a set of "parallel" time lines, all sharing the same structure of time. Each world describes a possible course of the universe over the same period of time. Hence, over the same period of time, but in different worlds, different facts are true. An example of such a NMEL-model is the following structure:

	t_1	t_2	t_3	t_4	t_5
w_1					
w_2					
w_3					
.					
w_n					

Here, w_i is a possible world, and t_i a timepoint. A NMEL-model has a S5-accessibility relation between the worlds, hence explicit mention of the accessibility relation is not necessary. The operators \Box and \Diamond are defined in NMEL as follows:

$$\begin{aligned} M \models \Box \varphi &\Leftrightarrow \text{for all } w \text{ in } M: M, w \models \varphi \\ M \models \Diamond \varphi &\Leftrightarrow \text{for at least one } w \text{ in } M: M, w \models \varphi \end{aligned}$$

Here $M, w \models \varphi$ means that φ is true in world w in model M . To illustrate the difference between NMEL and a monotonic epistemic logic, consider the following example:

$$\begin{array}{l} \text{(III)} \quad (1) \ B \\ \quad \quad (2) \ B \wedge \Diamond C \rightarrow D \\ \hline \quad \quad \quad D \end{array}$$

This argument is not valid in a monotonic logic, because the premise $\Diamond C$ is lacking. In NMEL however this argument is valid. In other words one can prove that the following holds:

$$B, (B \wedge \Diamond C) \rightarrow D \vDash_C D.$$

This can be argued for as follows. $B, (B \wedge \Diamond C) \rightarrow D \vDash_C D$ is valid, if D is true in all minimal models of the set of premises $(B, (B \wedge \Diamond C) \rightarrow D)$. It is simple to prove that this is the case. We will give a sketch of the argument. Consider a minimal model M of this set of premises, i.e., $M \vDash_C B \wedge (B \wedge \Diamond C) \rightarrow D$. This implies that

$$(1) \quad M \models B \text{ and } M \models (B \wedge \Diamond C) \rightarrow D.$$

As $\Box \neg C$ is not a classical S5 deductive consequence of this set of premises, this formula will not be satisfied by the minimal model M , i.e., $M \not\models \Box \neg C$.¹⁰ Consequently, we have $M \models \neg \Box \neg C$. So, by definition, it follows that

$$(2) \quad M \models \Diamond C.$$

From 1 and 2 it follows that $M \vDash D$, which finishes the sketch of the argument.

Note that if a model is minimized with respect to the number of \Box -formulas, then it is in a sense *maximized* with respect to the number of \Diamond -formulas. Because if M does not satisfy $\Box \neg \varphi$, then it does satisfy $\neg \Box \neg \varphi$, and hence by definition it does satisfy $\Diamond \varphi$. This corresponds to the intuitive reading of the \Diamond -operator, which says that one can assume φ , unless one already knows that $\neg \varphi$ is the case.

After this short introduction to NMEL, we can explain how NMEL can be used to formalize P's explanation of the burning of the lamp in the electric circuit. The problem is that although P is unable to predict the burning of the lamp, he can very well explain afterward why the lamp went on at t_2 . This asymmetry of prediction and explanation is due to the fact that P, not having complete knowledge, is unable to use the empirical law W. However in his explanation P is supposed to arrive at an empirical law W'; which is a close approximation of W. First, we will present the NMEL analysis of P's explanation. After that we will show how NMEL accounts for the fact that P is unable to predict the burning of the lamp.

At timepoint t_2 P is pressing switch S_1 , and subsequently he observes that the lamp L goes on. Hence, P knows that at t_2 both S_1 and L are on, which is expressed by $\Box(2, S_1 = \text{on})$ and $\Box(2, L = \text{on})$. These observations led P to the conjecture that the burning of the light was caused by switching S_1 . This explanatory argument can be formalized in NMEL as follows:

- (IV) (1) $\Box(2, S_1 = \text{on})$
 (2) $\Box(2, L = \text{on})$

$$\Diamond \forall t [((t, S_1 = \text{on}) \wedge \Diamond(t, \neg D)) \rightarrow (t, L = \text{on})] \quad (C_1)$$

This argument is analogous to argument (II').¹¹ It expresses that, given the knowledge P has about the circuit at t_2 , he can assume that the law $\forall t [((t, S_1 = \text{on}) \wedge \Diamond(t, \neg D)) \rightarrow (t, L = \text{on})]$ holds. This law says that turning on the switch will cause the lamp to burn, provided one can assume that there are no distorting factors in the circuit. This argument is valid in NMEL, and not valid in a monotonic epistemic logic.¹² For proofs of this and subsequent arguments the reader is referred to the appendix.

The NMEL analysis also accounts for the fact that P cannot ex ante

predict at t_1 that the light will go on, when S_1 is turned on at the next moment t_2 . At t_1 P only knows that he will turn on the switch S_1 at the next moment t_2 . As P is not even aware of the existence of the other switch S_2 , he certainly doesn't know that this switch is on at t_2 , and neither does he know the law W, i.e. $\forall t[(t, S_1 = \text{on}) \wedge (t, S_2 = \text{on})] \rightarrow (t, L = \text{on})$. Thus, at t_1 P may argue as follows:

$$\begin{array}{l} \text{(V)} \quad (1) \quad \Box(2, S_1 = \text{on}) \\ \hline \quad \quad \quad \neg\Box(2, L = \text{on}) \end{array}$$

Hence, although P intends to turn S_1 on at t_2 , he does not know that the lamp will go on. This means that P cannot predict whether the light will be on at t_2 or not. This argument is valid in NMEL, and not valid in a monotonic epistemic logic.¹³

However, if P had complete knowledge about the circuit, he could predict that the light will go on. This can be argued for as follows. At t_1 P intends to press S_1 at t_2 , so he knows that this switch will be on at t_2 ; i.e., $\Box(2, S_1 = \text{on})$. Furthermore, having complete knowledge, P knows the law W, i.e., $\Box\forall t[(t, S_1 = \text{on}) \wedge (t, S_2 = \text{on})] \rightarrow (t, L = \text{on})$, and he knows that the other switch S_2 will still be on at t_2 , i.e., $\Box(2, S_2 = \text{on})$. Thus, at t_1 P could argue as follows:

$$\begin{array}{l} \text{(VI)} \quad (1) \quad \Box(2, S_1 = \text{on}) \\ \quad \quad (2) \quad \Box(2, S_2 = \text{on}) \\ \quad \quad (3) \quad \Box\forall t[(t, S_1 = \text{on}) \wedge (t, S_2 = \text{on})] \rightarrow (t, L = \text{on}) \\ \hline \quad \quad \quad \Box(2, L = \text{on}) \end{array}$$

This predictive argument is valid in NMEL. It is also valid in a monotonic epistemic logic. Hence, the validity of this argument does not depend on the non-monotonicity property of NMEL. The reason that we mention this example is that it shows that Hempel's covering law model is simply a special (ideal!) case in NMEL. If P has complete knowledge, the NMEL analysis of predictions is analogous to Hempel's analysis.

With respect to P's explanation, i.e., argument IV, we have to make one final comment. The following explanatory argument is also valid in NMEL:

$$\begin{array}{l}
 \text{(IV')} \quad (1) \quad \Box(2, S_1 = \text{on}) \\
 \quad \quad (2) \quad \Box(2, L = \text{on}) \\
 \hline
 \quad \quad \Diamond \forall t [((t, L = \text{on}) \wedge \Diamond(t, \neg D)) \\
 \quad \quad \rightarrow (t, S_1 = \text{on})] \quad (C_2)
 \end{array}$$

The difference with (IV) is that in the conclusion of this argument the causal direction is reversed; C_2 says that S_1 being on is caused by the light going on. We don't think NMEL is to be blamed for this ambiguity of causal direction.¹⁴ It is always hard to establish the causal direction between two (almost) simultaneous events. It only indicates that in explanatory arguments it is not only logic that matters. Pragmatic considerations also have to be taken into account. Actually, this problem was already discussed in Section 2. There it was observed that the 'Powerful-God' explanation, which is analogous to C_2 , was excluded by general background knowledge about electricity. Although our NMEL analysis does not exclude the possibility that P could have argued as in (IV'), this argument is as it were overruled by general background knowledge. Technically, this result could easily be obtained by extending the set of premises in (IV) with the set T, that includes all the knowledge sentences representing the general background knowledge. Presupposing that T does indeed exclude C_2 , i.e., $T \vDash \neg C_2$, and that T does not exclude C_1 , i.e., $T \not\vDash \neg C_1$, we would get the following argument in NMEL:

$$\begin{array}{l}
 \text{(VII)} \quad (1) \quad \Box(2, S_1 = \text{on}) \\
 \quad \quad (2) \quad \Box(2, L = \text{on}) \\
 \quad \quad (3) \quad T \\
 \hline
 \quad \quad C_1 \wedge \neg C_2
 \end{array}$$

Hence, if we would really take all of P's general background knowledge into account, we arrive at exactly the intended explanatory argument. The problem with T is of course that usually this background knowledge is so extensive that we won't be able to write it down. Consequently, the role of background knowledge is inherently obscure in every logical analysis of explanatory arguments. It is for this reason that we consider background knowledge to be a pragmatic aspect of explanatory arguments.

5. A LOGICAL RECONSTRUCTION OF FRIEDMAN'S MONETARY HISTORY

In this section we attempt to reconstruct Friedman's arguments in terms of NMEL. We focus attention upon the unspecified *ceteris paribus* clause Friedman uses. In the first part of 'Money and Business Cycles' Friedman describes some facts about the rate of change of the money stock ($\dot{M}_t = \alpha_t$) and of nominal income ($\dot{Y}_t = \beta_t$). It is remarked that turning points in the rate of change of the money stock show a lead over turning points in the rate of change of income and this relation is almost constant (c) over a long period that has seen substantial changes. He concludes that a relation may exist between the variability of the stock of money and the variability of income. This relation would probably have been a perfect correlation provided that there were no distorting factors ($\neg D_S$) with respect to the system (S) under consideration. Formally we may represent the argument as follows;

$$\begin{array}{ll} \square(t, \dot{M}_t = \alpha_t) & \text{for } t = 1867, \dots, 1960 \\ \square(t, \dot{Y}_t = \beta_t) & \text{for } t = 1867, \dots, 1960 \\ \square(t, \alpha_{t-i} \approx c, \beta_t) & \text{for } t = 1867, \dots, 1960 \end{array}$$

-
- (i) $\diamond \forall t: [((t, \dot{M}_t = \alpha_t) \wedge \diamond(t, \neg D_S)) \rightarrow (t, \dot{Y}_t = \beta_t)]$ (and also)
(ii) $\diamond \forall t: [((t, \dot{Y}_t = \beta_t) \wedge \diamond(t, \neg D_S)) \rightarrow (t, \dot{M}_t = \alpha_t)]$

This scheme resembles the structure of P's explanation with respect to the electric circuit example; see scheme IV in Section 4. Remark that the 'assume' operator $\diamond X$ does precisely what Friedman would have had in mind: we do not know X, but as we have no contradictory evidence (partly because of a lack of sufficient detailed knowledge) we may assume X to be the case.

In the first part of the argument Friedman does not make a choice between the two possible conclusions (i) and (ii). Eventually, however, he has to make a choice because of the intuitive idea that causality runs only in one direction. It is interesting to notice that a condition as "the cause comes first" (see e.g., Nagel 1961, p. 74) is not conclusive in this context, because (as Figure 3 shows) one cannot be sure which one of two cycles comes first: apparent leads may "really" be lagged responses. These difficulties are further complicated, because anticipations play a crucial role in economics. The argument of the second part allows for another use of NMEL. Here, Friedman knows by definition ("arith-

metically”) that changes in the money stock can be attributed to three proximate determinants. For each of the major changes in history he shows that one of these (autonomous) factors is present. Thereby, he frankly admits that “historically there have been many different reasons” (p. 221). Especially for the minor changes in the stock of money he explicitly adheres to an incompletely specified *ceteris paribus* clause (see Section 3). If we use AHF for “autonomous historical factors” the argument can be represented in the following formal way.

$$\begin{array}{l} \square \forall t: [((t, \text{AHF}) \wedge \diamond(t, \neg D_{S3})) \rightarrow (t, M_t = \alpha_t)] \\ \square(t', \text{AHF}) \\ \hline \square(t', M_t = \alpha_t) \end{array}$$

(t' denotes a constant, i.e., a specific timepoint, and t denotes a variable.) This scheme is similar to scheme III in Section 4. Also this argument is not valid in a standard logical account of explanation, because in such an account all potentially distorting factors have to be mentioned explicitly and, hence, have to be known. On the basis of the above argument Friedman poses that the direction of causality runs from money to income. So, he narrows down the set of admissible conclusions in the first part of the argument. In other words: the second part serves as background knowledge which eliminates conclusion (ii) of the first part.

The third part of ‘Money and Business Cycles’ does not comprise a genuine argument. In an informal discussion it is demonstrated that the covariance between money and income is in consonance with (can be derived from) the accepted body of economic theory. For this purpose Friedman introduces again an unspecified *ceteris paribus* clause (see Section 3). The third part of the argument is thus a means to demonstrate that the proposed direction of causality is consistent with the economic theory contained in the background knowledge. Of course, Friedman does not show that economic theory excludes the possibility that causality runs from income to money (cf. Section 4).

The crucial role of the background knowledge is nicely illustrated by the debate between Friedman and one of his critics Tobin. In Tobin (1971) Friedman is criticized for making the *post hoc ergo propter hoc* fallacy; i.e., concluding that the event B is caused by event A, simply because B is preceded by A. Friedman anticipates this criticism in his paper ‘Money and Business Cycles’. Suppose, he says, that we have a

similar body of evidence about the relation between the production of dressmakers' pins and income as we have about the relation between money and income, i.e., an increase in pin production is usually followed by an increase in income. This evidence, Friedman remarks, "would persuade neither us nor our readers to adopt a pin theory of business cycles" (p. 214), i.e., that an increase in income would be caused by an increase in pin production. The question is why we decide in favour of a monetary theory of business cycles? This choice is motivated by Friedman by referring to supplementary evidence. This supplementary evidence is what we have called background knowledge and is provided in the second and third part of 'Money and Business Cycles'.¹⁵

6. CONCLUSIONS

In this paper we have shown that explanations based on incomplete information do not always comply with Hempel's covering law model. In the example of the 'hidden' circuit we have seen that the relevant specific law (W) was not known beforehand, hence prediction was impossible. Still, person P could explain with hindsight why the lamp went on. Somehow person P was capable of inferring a law (W'), which is an approximation of W, from the observations that the switch S₁ and the lamp went on at the same time. To account for this type of reasoning we introduced scheme II as a supplement to Hempel's covering law model. And we discussed a logic (NMEL) which validates arguments like scheme II. Furthermore it was shown that our NMEL analysis does not only apply to very simple explanations, but it is applicable to Friedman's sophisticated explanation of the monetary history of the U.S. as well. This is an interesting test case, because it has been very much disputed. Friedman's analysis is neither a simple case of a covering law explanation, as we pointed out, nor is it a straightforward application of his own 'as-if' methodology. NMEL gives a fresh look on Friedman's actual line of reasoning in 'Money and Business Cycles'. Different aspects of his (and Tobin's) arguments can all be formalized within NMEL. So, if our claims are correct, then the NMEL-analysis gives a more adequate account of causal explanations with incomplete knowledge than Hempel's covering law model.

Needless to say that our analysis of explanation is of a tentative character. We are still somewhat surprised ourselves that our analysis of explanation yields results that differ so much from Hempel's analysis.

If our critical comments are correct in this respect, how does it come that Hempel's covering law model has survived so many decades? A possible answer to this question could be that philosophy of science *itself* is dominated by (logical) paradigms. As long as classical logic was the only logic available, people tend to lay every methodological issue on the procrustean bed of classical logic. Now that we have alternative logics at our disposal we can have a fresh look at these matters. Hence, non-monotonic logic might provide a new paradigm for the logical analysis of science. But of course putting a Kuhnian label on the NMEL-analysis doesn't justify it. In order to give a firmer base to our NMEL-analysis, we have to study to what extent it can be generalized. With this goal in mind there are at least two issues that have to be further investigated in subsequent research. First, we plan to write an article about the relation between Friedman's 'as-if' methodology and non-monotonic logic. In the introduction it was said that debates about the question to what extent Friedman did apply his 'as-if' methodology to his own research have obscured the analysis of his actual research. However, we didn't deny that aspects of the 'as-if' methodology can be pointed out in his research. Although we didn't say much in this paper about the debate on Friedman's 'as-if' methodology, we have strong indications that what is called diagnostic reasoning in artificial intelligence can shed new light on this debate. As non-monotonic logic is an essential ingredient of diagnostic reasoning, it is to be expected that NMEL could be very useful for a better understanding of the 'as-if' methodology as well.¹⁶ Second, in the philosophy of science there is a long-standing tradition of critical comments on Hempel's covering law model. It is interesting to study to what extent our NMEL analysis can deal with the comments made by other authors.

APPENDIX

In this appendix we will give the syntax and semantics of NMEL. NMEL is a point-based temporal logic augmented by the modal operators \square and \diamond . NMEL is a simplified version of Shoham's non-monotonic epistemic logic CI as it is defined in Shoham (1988a, b). NMEL is simpler than CI, because NMEL is point-based, whereas Shoham's logic is internal-based. This simplification is not essential.

Syntax of NMEL

Given P : a set of primitive propositions, TV : a set of temporal variables, e.g., t , TC : the set of temporal constants $\{\dots, -2, -1, 0, 1, 2, \dots\} \cup \{t'\}$, $U: TC \cup TV$, and \leq a binary relation symbol, the set of *well-formed formulas* (wffs) of NMEL is defined inductively as follows:

- (1) If $u_1 \in U$ and $u_2 \in U$, then $u_1 = u_2$ and $u_1 \leq u_2$ are wffs.
- (2) If $u \in U$ and $p \in P$, then (u, p) is a wff.
- (3) If φ_1 and φ_2 are wffs, then so are $\varphi_1 \wedge \varphi_2$, $\neg\varphi_1$, $\Box\varphi_1$.
- (4) If φ is a wff and $t \in TV$, then $\forall t\varphi$ is also a wff.

We assume the usual definitions of \vee , \leftrightarrow , \exists , and so on. $\Diamond\varphi$ is defined by $\Diamond\varphi \leftrightarrow \neg\Box\neg\varphi$. With respect to negation we have $(u, \neg\varphi) \leftrightarrow \neg(u, \varphi)$. In expressions like $(1, S_1 = \text{on})$, $(1, M_t = \alpha_t)$ and (t, AHF) , the subexpressions $S_1 = \text{on}$, $M_t = \alpha_t$ and AHF are considered to be primitive propositions, to keep NMEL as simple as possible. The generalization to a full predicate logical version of NMEL, i.e., not just quantification over temporal variables but also over object variables, is straightforward.

Semantics of NMEL

We start with the following definitions. \mathbb{N} is used to denote the natural numbers with the standard ordering \geq . A *model* M is a pair $\langle W, I \rangle$ where W is a nonempty universe of possible worlds, and I is an interpretation function $I: P \rightarrow 2^{W \times \mathbb{N}}$.¹⁷ No accessibility relation (R) between worlds is mentioned, because this relation is supposed to be universal, i.e., $\forall w, w' \in W: wRw'$. For an example of such a model, see Section 4. A *variable assignment* is a function $VA: TV \rightarrow \mathbb{N}$. If $u \in U$ then we define $VAL(u)$ to be $VA(u)$ if $u \in TV$, and the standard interpretation of u if $u \in TC$. A formula φ is true in (or technically speaking *satisfied by*) a world w of a model M under the variable assignment VA , written $M, w \models [VA]$, under the following conditions:

- $$\begin{aligned} M, w \models u_1 = u_2[VA] &\text{ iff } VAL(u_1) = VAL(u_2) \\ M, w \models u_1 \leq u_2[VA] &\text{ iff } VAL(u_1) \leq VAL(u_2) \\ M, w \models (u, p)[VA] &\text{ iff } (w, VAL(u)) \in I(p) \\ M, w \models \varphi_1 \wedge \varphi_2[VA] &\text{ iff } M, w \models \varphi_1[VA] \text{ and } M, w \models \varphi_2[VA] \\ M, w \models \neg\varphi[VA] &\text{ iff } M, w \not\models \varphi[VA] \end{aligned}$$

$M, w \vDash \forall t\varphi[VA]$ iff $M, w \vDash \varphi[VA']$ for all alternative assignments VA' that agree with VA everywhere except possibly on t .

$M, w \vDash \Box\varphi[VA]$ iff $M, w' \vDash \varphi[VA]$ for all $w' \in W$

$M, w \vDash \Diamond\varphi[VA]$ iff $M, w' \vDash \varphi[VA]$ for at least one $w' \in W$
 (Therefore we will be able to write $M \vDash \Box\varphi[VA]$ and $M \not\vDash \Box\varphi[VA]$ without fear of ambiguity, and analogous for $\Diamond\varphi$.)

The reader familiar with modal logic will note that from the identity of time across worlds follows the validity of the "Barcan formula", i.e., $\Box\forall t\varphi \leftrightarrow \forall t\Box\varphi$.

With respect to the syntax and semantics defined above we can define two entailment relations; a monotonic and a non-monotonic one. The monotonic entailment relation yields the monotonic version of NMEL. This logic is called TK in Shoham (1988a, b). The non-monotonic entailment relation yields NMEL itself. This logic is called CI in Shoham (1988a, b).

DEFINITION 1. *Monotonic entailment*; Σ monotonically entails φ , written $\Sigma \vDash \varphi$, if for all models M and all worlds w in M :

$$\text{If } M, w \vDash \Sigma[VA] \text{ then } M, w \vDash \varphi[VA].$$

In other words, if Σ is true in a world w of a model M , then φ is true in this world w .

DEFINITION 2. A wff φ is *valid* if for all models M and all worlds w in M and all assignments VA :

$$M, w \vDash \varphi[VA].$$

DEFINITION 3. *Non-monotonic entailment*; Σ non-monotonically entails φ , written $\Sigma \vDash_C \varphi$, if for all *minimal* models M of Σ and all worlds w in M :

$$\text{If } M, w \vDash_C \Sigma[VA] \text{ then } M, w \vDash \varphi[VA].$$

In other words, if Σ is true in a world w of a minimal model M of Σ , then φ is true in this world w . In order to define the notion of minimality we first have to give some preliminary definitions.

DEFINITION 4. A wff is *base* if it does not contain any occurrence of the modal operators \Box or \Diamond .

DEFINITION 5. The latest timepoint (ltp) of a base formula is the latest timepoint mentioned in it.

- (1) The ltp of (t, p) is t .
- (2) The ltp of $\varphi_1 \wedge \varphi_2$ is the latest between the ltp of φ_1 and the ltp of φ_2 .
- (3) The ltp of $\neg\varphi$ is the ltp of φ .
- (4) The ltp of $\forall t\varphi$ is the earliest among the ltp's of all φ' which result from substituting in φ a time point for all free occurrences of t , or 0 if there is none.

DEFINITION 6. A model M_2 is *more ignorant* than a model M_1 , written $M_1 \subset M_2$, if there exists a time point t such that

- (1) for any base sentence φ whose ltp $\leq t$, if $M_2 \vDash \Box\varphi$ then also $M_1 \vDash \Box\varphi$, and
- (2) there exists some base sentence φ whose ltp is t such that $M_1 \vDash \Box\varphi$, but $M_2 \not\vDash \Box\varphi$.

If M_2 is more ignorant than M_1 , then M_2 satisfies less formulas of the form $\Box\varphi$ (with φ a base wff) than M_1 .¹⁸

DEFINITION 7. M is said to be a *minimal knowledge model* of φ , written $M \vDash_{\subset} \varphi$, if $M \vDash \varphi$ and there is no other M' such that $M' \vDash \varphi$ and $M \subset M'$.

We will now present some results that are relevant for the arguments in the paper. Subsequently we will present the proofs of some of the arguments in the paper.

As NMEL-models have an accessibility relation R that is universal, all the S5 valid formulas are valid in NMEL.¹⁹ Let α , β and γ be well-formed formulas, then the following S5-theorems hold in NMEL:

THEOREM 1. $M \vDash \Diamond \forall t[(t, \alpha) \rightarrow (t, \gamma)] \Rightarrow M \vDash \forall t[((t, \alpha) \wedge (t, \beta)) \rightarrow (t, \gamma)]$.

Proof. Trivial, use the S5 valid formulas $\varphi \rightarrow (\psi \rightarrow \varphi)$ and $(\varphi \rightarrow (\psi \rightarrow \varphi)) \leftrightarrow ((\varphi \wedge \psi) \rightarrow \varphi)$.

THEOREM 2. $M \vDash \Box \forall t[((t, \alpha) \wedge (t, \beta)) \rightarrow (t, \gamma)] \Rightarrow M \vDash \forall t[(\Box t, \alpha) \wedge \Box(t, \beta) \rightarrow \Box(t, \gamma)]$.

Proof. Trivial, use the Barcan formula $\Box\forall t\varphi \leftrightarrow \forall t\Box\varphi$ and the S5 valid formulas $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$ and $\Box(\varphi \wedge \psi) \leftrightarrow (\Box\varphi \wedge \Box\psi)$.

THEOREM 3. The (unique) set of base sentences that are known in any minimal knowledge model of a causal theory is *exactly* the set of all those formulas that are S5-entailed by all (positive and negative) atomic base sentences that are known in that minimal knowledge model.

Proof. See Corollary 4.5 in Shoham (1988b).²⁰

As Shoham's definition of the notion *causal* is rather complicated, we will omit it here. In the following proofs of the arguments Theorem 3 will only be applied if the premises of the argument constitute a causal theory. For example the premises in argument (VI) do not constitute a causal theory, because the third premise is not causal. However, the conclusion of this argument can be proved without Theorem 3.

Proof of argument (IV).

- (1) $\Box(2, S_1 = on)$
- (2) $\Box(2, L = on)$

$$\diamond\forall t[(t, S_1 = on) \wedge \diamond(t, \neg D)] \rightarrow (t, L = on) \quad (C_1)$$

Consider an arbitrary *minimal* model M which makes the premises 1 and 2 true. Due to the minimality of M it is obvious that 1 and 2 are the only atomic base formulas known in M. Furthermore, it is also obvious that the formula $\Box\neg\forall t[(t, S_1 = on) \rightarrow (t, L = on)]$ is not S5-entailed by the premises 1 and 2. Hence, due to the minimality of M and Theorem 3 it follows that

$$\begin{aligned} & M \models \neg\Box\neg\forall t[(t, S_1 = on) \rightarrow (t, L = on)] \\ \Leftrightarrow & M \models \diamond\forall t[(t, S_1 = on) \rightarrow (t, L = on)]. \end{aligned}$$

Hence, by Theorem 1,

$$\Rightarrow M \models \diamond\forall t[(t, S_1 = on) \wedge \diamond(t, \neg D)] \rightarrow (t, L = on).$$

Argument (IV) is not valid in a monotonic epistemic logic, i.e., Shoham's logic TK. This can be shown by the following counter example. Consider a model M' that satisfies the formulas $\Box(2, S_1 = on)$, $\Box(2, L = on)$, $\Box(3, S_1 = on)$ and $\Box(3, L = off)$. However, as M' is not minimal

with respect to the knowledge at t_2 , it is not a minimal model of the premises 1 and 2.

Proof of argument (V).

$$\begin{array}{l} (1) \quad \Box(2, S_1 = \text{on}) \\ \hline \neg \Box(2, L = \text{on}) \end{array}$$

Consider an arbitrary minimal model M of premise 1. Note that premise 1 is the only atomic base formula known in M . It is sufficient to observe that the formula $\Box(2, L = \text{on})$ is not S5-entailed by this premise. Hence, due to the minimality of M and Theorem 3, it follows that $M \models \neg \Box(2, L = \text{on})$.

Argument (V) is not valid in a monotonic epistemic logic, as the following counter example shows. Note that the following set of sentences is S5 consistent: $\{\Box(2, S_1 = \text{on}), \Box(2, L = \text{on})\}$. Hence, it has a model, say M' . M' is a counter example for the argument. However, as M' is not minimal with respect to the premise 1, it is not a minimal model of this premise.

Proof of argument (VI).

$$\begin{array}{l} (1) \quad \Box(2, S_1 = \text{on}) \\ (2) \quad \Box(2, S_2 = \text{on}) \\ (3) \quad \Box \forall t [((t, S_1 = \text{on}) \wedge (t, S_2 = \text{on})) \rightarrow (t, L = \text{on})] \\ \hline \Box(2, L = \text{on}) \end{array}$$

Consider an arbitrary minimal model M of the premises 1, 2 and 3. Due to Theorem 2, premise 3 implies $\forall t [(\Box(t, S_1 = \text{on}) \wedge \Box(t, S_2 = \text{on})) \rightarrow \Box(t, L = \text{on})]$. Consequently, the instantiation $[(\Box(2, S_1 = \text{on}) \wedge \Box(2, S_2 = \text{on})) \rightarrow \Box(2, L = \text{on})]$ is also true in M . And, together with the premises 1 and 2, this implies that $M \models \Box(2, L = \text{on})$.

(Note that the minimality of M is not used in the proof. Consequently, the proof also holds for monotonic epistemic logic, i.e., the logic TK.)

Proof of argument (IV').

- (1) $\Box(2, S_1 = \text{on})$
- (2) $\Box(2, L = \text{on})$

$$\Diamond \forall t [((t, L = \text{on}) \wedge \Diamond(t, \neg D)) \rightarrow (t, S_1 = \text{on})] \quad (C_2)$$

Analogous to the proof of (IV), observe that $\Box \neg \forall t [(t, L = \text{on}) \rightarrow (t, S_1 = \text{on})]$ is not S5-entailed by the premises 1 and 2.

Proof of argument (VII).

- (1) $\Box(2, S_1 = \text{on})$
- (2) $\Box(2, L = \text{on})$
- (3) T

$$C_1 \wedge \neg C_2$$

Consider an arbitrary minimal model M that satisfies the premises 1, 2 and T. It was presupposed that $\neg C_2$ is S5-entailed by 1, 2 and T, i.e., $1, 2, T \vDash \neg C_2$. Furthermore it is presupposed that 1, 2 and T do not exclude C_1 , i.e., $1, 2, T \not\vdash C_1$. Hence, $\neg C_1$ is not S5-entailed by 1, 2 and T. If we can prove that $M \vDash C_1$ and $M \vDash \neg C_2$, then it follows immediately that $M \vDash C_1 \wedge \neg C_2$. $M \vDash \neg C_2$ holds, because $M \vDash 1 \wedge 2 \wedge T$ and $1, 2, T \vDash \neg C_2$. The proof of $M \vDash C_1$ is more complicated. First observe that $1, 2, T \not\vdash \neg C_1$ means that $1, 2, T \not\vdash \Box \neg \forall t [((t, S_1 = \text{on}) \wedge \Diamond t, (t, \neg D)) \rightarrow (t, L = \text{on})]$. Hence, there is a model N such that $N \vDash 1 \wedge 2 \wedge T$ and $N \not\vdash \Box \neg \forall t [((t, S_1 = \text{on}) \wedge \Diamond(t, \neg D)) \rightarrow (t, L = \text{on})]$. It is simple to see that then also $N \not\vdash \Box \neg \forall t [(t, S_1 = \text{on}) \rightarrow (t, L = \text{on})]$. Consequently we also have $1, 2, T \not\vdash \Box \neg \forall t [(t, S_1 = \text{on}) \rightarrow (t, L = \text{on})]$. As this last formula is of the form $\Box \varphi$, with φ a base formula, we can apply Theorem 3 with the result that $M \vDash \neg \Box \neg \forall t [(t, S_1 = \text{on}) \rightarrow (t, L = \text{on})]$.²¹ This implies, due to Theorem 1 and the definition of \Diamond , that $M \vDash \Diamond t [((t, S_1 = \text{on}) \wedge \Diamond(t, \neg D)) \rightarrow (t, L = \text{on})]$. Hence, we have $M \vDash C_1$.

NOTES

* We gratefully acknowledge the helpful suggestions and critical comments of Johan van Benthem, Bert Hamminga, Kevin Hoover, Theo Kuipers, and Rick Looyen.

¹ See for example Hirsch and de Marchi (1986) and also Hammond (1988).

² See Hempel (1965).

³ For our view on the important role of the *unspecified* ceteris paribus clause we are very much indebted to Kuipers (1986). Furthermore, Kuipers was the one who suggested that we should use electric circuits to illustrate the role of incomplete knowledge in causal explanations. For a comparison between his ideas and ours, see Tan (1988).

⁴ As Friedman is usually regarded as the main author of this work, we will in the sequel simply refer to Friedman.

⁵ Naturally, there is a difference in degree, because electricity laws are better established than economic theoretical laws.

⁶ For a survey of non-monotonic logic see Ginsberg (1987).

⁷ Those readers familiar with Hintikka's epistemic logic (Hintikka 1962) will observe that \diamond is not the same as Hintikka's belief operator. From a technical point of view \diamond behaves as a possibility operator in modal logic. This has the effect that a formula of the form $\diamond\varphi \wedge \diamond\neg\varphi$ is *not* inconsistent in NMEL.

⁸ The fact that *no* formula of the form $\Box\varphi$, which is not a classical S5 deductive consequence of A, is satisfied by a minimal model M of A, is justified by Theorem 3 in the appendix. B is a classical S5 deductive consequence of A, if B is derivable from A in classical S5 modal logic, i.e., the logic TK in Shoham (1988a, b). For further details about TK see the appendix.

⁹ See Hughes and Cresswell (1972).

¹⁰ It is presupposed that B is not logically equivalent with $\Box\neg C$.

¹¹ The proof of (IV) indicates that C_1 is not the strongest conclusion entailed by the premises 1 and 2; the strongest conclusion is $\diamond\forall t[(t, S_1 = \text{on}) \rightarrow (t, L = \text{on})]$. However in (IV) we mention C_1 instead of this stronger conclusion to emphasize the analogy with (III).

¹² By a monotonic epistemic logic we mean a monotonic version of NMEL; i.e., a logic that is identical to NMEL except that the minimal entailment operator \vDash_{C} is replaced by a classical entailment operator \vDash . See the logic TK in the appendix.

¹³ This line of arguments shows that it is easier to explain than to predict. In predictions we have to know the relevant covering laws, whereas in explanations we can do without them.

¹⁴ Another objection could be that two opposite (although not contradictory!) conclusions, such as C1 and C2, both follow in NMEL from the very same premises. However, a closer look reveals that this happens in classical logic too. For example in classical propositional logic both $p \rightarrow q$ and $q \rightarrow p$ follow from the set of premises $\{p, q\}$.

¹⁵ Tobin (1971) is also interesting for another reason. He shows that the data mentioned by Friedman can also be accounted for by a theory that proposes exactly the opposite causal direction, i.e., an increase of money is caused by an increase in income. In order to make his case Tobin criticizes Friedman's claim that the factors determining the increase of money are autonomous (AHF). Tobin provides an alternative mechanism which explains changes in money in terms of changes in income. Hence, the debate between Tobin and Friedman is essentially about the correctness of Friedman's background knowledge.

¹⁶ For a detailed analysis of non-monotonic aspects in diagnostic reasoning see for example Reiter (1987). An interesting alternative approach to diagnostic reasoning is proposed in Jackson (1988).

¹⁷ In Shoham (1988a, b) the range of the interpretation function I is $2^{W \times N \times N}$ instead of

$2^{W \times N}$. The reason is that NMEL is point based, whereas Shoham's logic CI is interval-based. This simplification is not essential.

¹⁸ Note that 'C' is not interpreted as set-inclusion.

¹⁹ For further details about S5 modal logic see Hughes and Cresswell (1972).

²⁰ Corollary 4.5 says: "The (unique) set of base sentences that are known in any c.m.i. model of a causal theory is exactly the set of tautological consequences of all (positive and negative) atomic base sentences that are known in that c.m.i. model." C.m.i. models are what we call minimal knowledge models. The notion 'tautological consequence' is not explicitly defined by Shoham. However, from the definition of the semantics and Definition 1 it follows immediately that tautological consequence is equivalent to S_5 -entailed.

²¹ For the sake of the argument we will assume that T is a causal theory.

REFERENCES

- Blaug, M.: 1976, 'Kuhn versus Lakatos or Paradigms versus Research Programmes in the History of Economics', in S. Latsis (ed.), *Method and Appraisal in Economics*, Cambridge University Press, Cambridge.
- Cartwright, N.: 1983, *How the Laws of Physics Lie*, Oxford, Clarendon Press.
- Friedman, M.: 1953, 'The Methodology of Positive Economics' in *Essays in Positive Economics*, Chicago University Press, Chicago.
- Friedman, M.: 1969, 'Money and Business Cycles', in *The Optimum Quantity of Money and Other Essays*, Macmillan, London.
- Friedman, M. and A. Schwartz: 1963, *A Monetary History of the United States 1867-1960*, Princeton University Press, Princeton.
- Ginsberg, M. L. (ed.): 1987, *Readings in Non-Monotonic Logic*, Kaufmann, Los Altos.
- Hammond, J.: 1988, 'How Different are Friedman and Hicks on Method', *Oxford Economic Papers* 40, pp. 392-94.
- Hempel, C. G.: 1965, 'Scientific Explanation', in C. G. Hempel (ed.), *Aspects of Scientific Explanation*, Free Press, New York.
- Hintikka, J.: 1962, *Knowledge and Belief*, Cornell University Press, Cornell, New York.
- Hirsch, A. and N. de Marchi: 1986, 'Making A Case When Theory is Unfalsifiable: Friedman's Monetary History', *Economics and Philosophy* 2, 1-21.
- Hughes, G. E. and M. J. Cresswell: 1972, *An Introduction to Modal Logic*, Methuen and Co., London.
- Jackson, P.: 1988, *Diagnosis, Defaults and Abduction*, McDonell Douglas Research Laboratories, unpublished.
- Kuipers, T.: 1986, 'Explanation by Specification', *Logique et Analyse* 116, 509-21.
- Nagel, E.: 1961, *The Structure of Science*, Routledge and Kegan Paul, London.
- Reiter, R.: 1987, 'A Theory of Diagnosis from First Principles', *Artificial Intelligence*, 32, 57-96.
- Salmon, W.: 1989: 'Four Decades of Scientific Explanation', *Minnesota Studies in the Philosophy of Science*, vol. 13.
- Shoham, Y.: 1988a, *Reasoning about Change, Time and Causation from the Standpoint of Artificial Intelligence*, MIT Press, Boston.
- Shoham, Y.: 1988b, 'Chronological Ignorance, Experiments in Non-monotonic Temporal Reasoning', *Artificial Intelligence* 36, 279-331.

- Tan, Y. H.: 1988, 'Explanations with Incomplete Information; A Problem for Hempel's Theory about Causal Explanation' (in Dutch), in W. Callebaut and P. Mostert (eds.), *Congresbundel Filosofiedag Maastricht 1987*, Eburon, Delft.
- Tobin, J.: 1971, 'Money and Income: Post Hoc Ergo Propter Hoc', *Essays in Macroeconomics*, University of Chicago Press, Chicago, pp. 497-514.

M. C. W. Janssen
Dept. of Philosophy
Erasmus University Rotterdam
P.O. Box 1738
3000 DR Rotterdam
The Netherlands

and

Y.-H. Tan
Dept. of Mathematics and Computer Science
Free University Amsterdam
de Boelelaan 1081
1081 HV Amsterdam
The Netherlands