Logique & Analyse **133-134** (1991), 197-220

NON-MONOTONIC EPISTEMIC ASPECTS OF SCIENTIFIC EXPLANATIONS

Yao-Hua TAN

Abstract

In this paper we show that explanations based on incomplete information do not always comply with Hempel's *Covering Law Model* of scientific explanation. We show that the relevant covering law is usually not known beforehand in this type of explanations. This results in a breakdown of the symmetry between prediction and explanation in Hempel's *DN*-model in the case of explanations based on incomplete information. We argue that in this type of explanations it is possible to derive from the observed facts a weaker type of law, which we call *unspecific laws*, which are strong enough for explanation, but too weak for prediction. Furthermore, we argue that this derivation of unspecific laws from observed facts presupposes a new type of arguments, which we called *Law-Finding-From-Facts* or *L3F* Arguments, which are supplementary to Hempel's covering law model. These *L3F* arguments cannot be modelled in classical logic, and should not be considered as inductive arguments either. We show that Shoham's non-monotonic epistemic logic is the best logic to model these *L3F* arguments.

1. *Introduction*

In recent years several non-monotonic epistemic logics have been developed. The best known ones are autoepistemic logic which was developed by Moore in [Moo85] and Konolige in [Kon89], and the non-monotonic epistemic logic that was developed by Shoham in [Sho88a, b]. In this paper we will not introduce yet another new non-monotonic epistemic logic, but instead we will advocate a new application of the existing non-monotonic epistemic logics, in particular the logic of Shoham. This new application is the modelling of explanations based on incomplete information.

From the sixties on numerous philosophers of science, in particular logical positivists like Carnap and Hempel, have tried to develop a logical model for scientific explanation (for an extensive survey of this research tradition see [Ste83] and [Sal90]). The most influential model of scientific explanation is Hempel's so-called *Deductive-Nomological Model*, or short *DN-Model* also called *Covering Law Model*, that he introduced in his famous article *Aspects of Scientific Explanation* (see [Hem65])([1]). For several decades, philosophers of science considered the covering law model basically correct. However, in recent years Hempel's model of explanation has been severely criticized (see e.g. [Sal90]). In this paper we add yet another critical comment on the covering law model; in particular we will argue that the covering law model is not adequate to deal with situations in which there is *incomplete information*. The covering law model presupposes complete information. It can only be applied to cases where everything is known: the relevant empirical law, initial conditions and absence of potential distorting factors. However in science these conditions are seldom met. In this paper we focus attention upon situations in which the relevant covering law is not known. We will show that in the case of incomplete knowledge intuitive appealing explanations are still feasible. However, in order to model these explanations based on incomplete information we had to introduce a new type of arguments, the so-called *Law-Finding-From-Facts* (*L3F*) arguments, of which we claim that they are different not only from any classically deductive argument but also from any other non-deductive argument such as inductive or abductive arguments. To clarify the logical structure of these *L3F* arguments we give a formalization of this type of explanation in non-monotonic logic, which is supplementary to the deductive scheme of Hempel's covering law model. We will show that in particular the non-monotonic epistemic logic (*NMEL*) that was introduced by Shoham is very suitable to formalize explanations based on incomplete knowledge. This paper is a further development of earlier research that was reported in [Tan88] and [JT91].

This paper is organised as follows. In Section 2 we briefly discuss the covering law model of scientific explanation, and we present an example of an electric circuit in which a large part of the circuit is hidden, and hence unknown to the observer. The example serves a dual purpose. It elucidates the concept of explanation based on incomplete information and the role of *L3F* arguments in these explanations. Moreover, some more substantial points about explanation and prediction based on incomplete knowledge can be learned from this electric circuit example. Among other things it is

[1] In this paper we will use the terms *DN-model* and *Covering Law Model* interchangeably.

shown that there is an asymmetry between explanation and prediction, which can not be accounted for in the covering law model. In Section 3 we present Shoham's non-monotonic logic *NMEL*. And in Section 4 we show how L3F arguments can be very neatly modelled in *NMEL*, and we discuss 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. Finally, in Section 5 some results of this paper will be summarized.

2. *Explanations Based on Incomplete Information*

In [Tan88] an example of an explanation with incomplete information is discussed which is not covered by Hempel's *Covering Law Model* of causal explanations. The essence of this example is that, due to the incompleteness of the available information the relevant law is not fully specified. In this section we will study this counter-example more closely.

2.1 *Hempel's DN-Model of Deductive Explanation*

According to Hempel an explanation is a deductive inference which can be roughly presented as follows:

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

$Q(a)$ 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 *law*, which constitutes the core of the deductive explanation. Such 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 the 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.

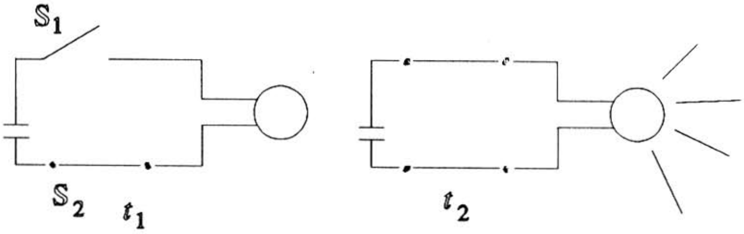

Fig. 1 Electric circuit at t_1 and t_2

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.

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

Substituting this law for the second premise in Hempel's *DN*-model, it is clear that the burning of the lamp is deductively explicable. 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.

2.2 *Two-Step Refinement of Hempel's DN-Model*

The real difficulty of an explanation is *not* to perform the deduction, but to *find* the specific empirical law that does the job. Try to remember how you solved physics problems at secondary school. The real difficulty was usually not how to calculate the solution, but to figure out how to apply the few fundamental laws that you had memorized to the specific data in the excercise. The basic problem was to make the appropriate derivations from the basic laws that would yield an empirical law which is specific enough to be directly applicable to the actual situation given in the excercise. After you made the appropriate derivations you just plugged in the figures, and the remaining calculation was usually routine.

According to Hempel this "law-finding" of the specific law that applies to a particular situation also has a *DN*-structure. He claims that a specific law can be deductively derived from the general background theory and the description of that actual situation. Let T denote in the sequel the *General Background Theory*, and let DA denote the *Description of the Actual situation*. For example, in the case of the electric circuit the background theory T is the theory about electricity, and the DA of this circuit is a description of the constituent components such as the two switches S_1 and S_2, the power supply, the lamp and the wires that connect these components etcetera. We could refine Hempel's *DN*-model into a so-called *Two-Step DN-Model*. In this two-step model we distinguish two consecutive steps. In the first step, the so-called *law-finding step*, we try to find a law which is specific enough to be applicable to the specific data of the actual situation we want to explain via a deduction from the relevant T and DA. In the second step, the so-called *law-applying step*, this specific law is actually applied to these specific data.

Two-Step DN-Model

1. *Law-Finding Step*:

$$\begin{array}{l} DA \\ T \\ \hline \forall x\ (P(x) \rightarrow Q(x)) \end{array}$$

2. *Law-Application Step*:

$$\begin{array}{l} P(a) \\ \forall x\ (P(x) \rightarrow Q(x)) \\ \hline Q(a) \end{array}$$

With this two-step model we can better pin down those aspects of explanation that make it problematic. Step 1 is the real challenge, whereas step 2 is a simple logical excercise. In the next section we will see that step 2 is especially problematic in the case of explanation based on incomplete information.

2.3 *Explanations based on incomplete information*

Our claim is that there are cases in which the symmetry in the *DN*-model breaks down. We illustrate this claim with the following counter example.

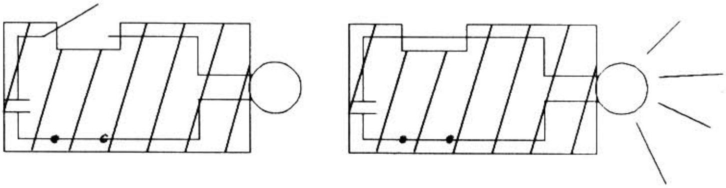

Fig. 2 Hidden electric circuit at t_1 and t_2

The situation in Figure 2 is similar to Figure 1, except that now the circuit is hidden 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 beforehand, 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 knowledge of 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 *DN*-model, 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 fully specified laws. Due to the lack of information 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, but P is not even aware of the existence of this second switch. However, P could come up with a close *approximation* U of W, i.e. something like

$$\forall t\ [((t,\ S_1 = \text{on}) \wedge (t,\ \neg DF)) \rightarrow (t,\ L = \text{on})] \qquad (U)$$

The expression $\neg DF$, which stands for "*there are no distorting factors in the system under consideration*", is a kind of *unspecified ceteris paribus clause*. We will call a law that contains such an unspecified ceteris paribus clause an *unspecified law*. 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, or that simply another switch in the circuit is open. [2]It is only when this ceteris paribus clause holds, that pressing S_1 causes the lamp to burn. Thus we arrive at an argument of the following type:

I. 1) $(2, S_1 = \text{on})$
2) $(2, L = \text{on})$

....

$\forall t\, [(t, S_1 = \text{on}) \wedge (t, \neg DF) \rightarrow (t, L = \text{on})]$

(The dotted line indicates that this is not a classically valid argument.)

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. Note that in this argument the unspecified law is, so to say, 'derived' from observed facts.

Our analysis of this example is the following. If possible we use the *DN*-model of law-finding from general theories to derive from the background theory *T* and the description of the actual situation *DA* the specific law *W*, i.e. we have the following law-finding derivation.

II. DA
T

$$\frac{\begin{array}{l}\text{DA}\\ \text{T}\end{array}}{\forall t\, [((t, S_1 = \text{on}) \wedge (t, S_2 = \text{on})) \rightarrow (t, L = \text{on})]}$$

However, if we have incomplete information about the actual situation, i.e. its description *DA* is incomplete, then this derivation of specific empirical laws becomes problematic. Instead we have in explanations based on incomplete information law-finding from observed facts, i.e. argument I. Note that argument I is almost the converse of argument II, i.e. the *DN*-model of law-finding as given in the two-step *DN*-model. Instead of deriving a specific

[2] The important role of the *unspecified* ceteris paribus clause in scientific explanations is also discussed in [Kui86]. For a comparison between his ideas and ours see [Tan88].

law from a more general theory T (combined with the DA), we derive the law from much more specific information; namely the observed facts! Moreover the general background theory T is not used as premise in this argument. We will later explain the role of T.

This argument above suggests a new type of argument of which the general form could be written as follows. For obvious reasons we call this type of argument law-finding-from-facts argument:

Law-Finding-From-Facts Argument:

1) $Q(a)$
2) $R(a)$

....

$\forall x\ ((Q(x) \wedge \neg DF(x)) \rightarrow R(x))$

In the sequel we will usually abbreviate *Law-Finding-From-Facts* arguments to *L3F* arguments. The basis idea of an *L3F* argument is that a specific law is 'inferred' from observed facts. Prima facie this argument scheme looks rather curious. It is a non sequitur in classical logic, as is indicated by the dotted line. This paradox will be solved in Section 3. We will see that the non-monotonic logic *NMEL* is an excellent logic to model *L3F* arguments.

Although the general background theory T is not used as a premise in the *L3F* argument, it does play a very important role in assessing the *adequacy* of an *L3F* argument. In the case of the hidden electric circuit the general background T about electricity does play a crucial role in P's explanation. If we accept the *L3F* argument given above, then there is prima facie no reason why we should not accept an analogous *L3F* argument with the causal direction reversed, i.e. concluding the law $\forall x\ ((R(x) \wedge \neg DF(x)) \rightarrow Q(x))$ from the premises $Q(a)$ and $R(a)$, which would give the following *L3F* argument.

1) $Q(a)$
2) $R(a)$

....

$\forall x\ ((R(x) \wedge \neg DF(x)) \rightarrow Q(x))$

In the case of the hidden electric circuit this would mean that we have the following pair of *L3F* arguments. Here *CU* denotes the following unspecifiec law

$$\forall t\ [((t, L = \text{on}) \wedge (t, \neg DF)) \rightarrow (t, S_1 = \text{on})] \qquad (CU)$$

with the reversed causality. The pressing of the switch S_1 is caused by the fact that the lamp went on.

III.1	$(t_2, S_1 = \text{on})$ $(t_2, L = \text{on})$ .... U	III.2	$(t_2, S_1 = \text{on})$ $(t_2, L = \text{on})$ .... CU

This law *CU* is not as weird as it might look like. Since 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. One can imagine that if *P* has no knowledge about electricty at all that, given the observed facts, he might have jumped to the reversed causal law *CU*. For example, 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. Hence, the general background knowledge *T* is used to *narrow down* the set of laws that could account for the burning of the light; i.e. it acts like a set of very strong 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. 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 pruning away by the general background theory *T* of unintended unspecific laws can be formally described as follows. The law *CU* is excluded by general theory *T*, whereas *U* is not excluded by *T*. We model this exclusion of *CU* with the classical entailment of the negation of *CU*. In other words, *CU* is excluded by a general theory *T* if *T* classically entails the negation of *CU*; i.e. $T \models \neg CU$. On the other hand *U* is not excluded by T, hence $T \not\models \neg U$. If we add these classical entailments to the *L3F* arguments as given above we expect, assuming that classical derivations are preserved by *L3F* arguments, the following pair of *L3F* arguments to hold:

$$\begin{array}{llll}
\text{IV.1} & T & \text{IV.2} & T \\
 & (t_2, S_1 = \text{on}) & & (t_2, S_1 = \text{on}) \\
 & (t_2, L = \text{on}) & & (t_2, L = \text{on}) \\
 & \cdots\cdots\cdots\cdots & & \cdots\cdots\cdots\cdots \\
 & U & & \neg CU
\end{array}$$

Argument IV.2 is also classically valid, whereas the argument IV.1 is not. However, in the next section we will see that this *L3F* argument is valid in *NMEL*. Generally speaking, we can say that there is a very subtle balance between *L3F* arguments and their relevant general background theory. The *L3F* arguments usually generate too many unintended laws, which are pruned away by the relevant background theory T. Perhaps one could even say that without knowledge of such a background theory *L3F* arguments are often non-sensical.

The morale of our analysis of the electric circuit example can be summarized 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. However, this simple deductive account of finding a specific law usually fails for explanations based on incomplete information. Since 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. The *L3F* argument provides the best guess, the unspecified law U, that P can come up with. Hence, *L3F* arguments are supplementary to Hempel's deductive account of explanation in the sense that it applies to cases with incomplete information which cannot be dealt with in Hempel's *DN*-model. One could perhaps argue that our hidden electric circuit example is too simple to justify such claims. However, In [JT91] we showed that the same analysis of explanations based on incomplete information holds for much more complicated explanations. We actually showed that Friedman's well-known explanation of the monetary history of the U. S. A. (see [Fri69]) can be analysed as an explanation based on incomplete information in which *L3F* arguments play a very prominent role.

2.4 *Law-finding-from-facts versus indiction or abduction*

In existing AI research there are currently two types of reasoning that are studied which differ from deductive reasoning: *inductive* reasoning and *abductive* reasoning (for further references see e.g. [LZSB87], [LZ89], [SS86], [Tan90] and [Fla92]). Although there are some obvious similarities between *L3F* arguments and these other types of non-deductive reasoning, we think it is substantially different from each of these types of non-deductive reasoning.

The similarity between Law-Finding-From-Facts and inductive reasoning is that in both cases laws are inferred from observed facts. However, in inductive reasoning the main objective is to *discover* and *corroborate* new empirical laws, whereas the main objective in *L3F* arguments is to *approximate* well-established laws. In *L3F* arguments there is absolutly no claim to novelty. If we use the unspecified law U to explain what happend in the hidden circuit, there is no claim that we have discovered a new empirical law about electricity. It is only that due to our lack of knowledge we have to jump, or perhaps it is better to say that we stumble, to our best guess. Furthermore, we pointed out that *L3F* arguments *only* make sense against the background of a well-established general background theory T. Without the boundary conditions provided by T, *L3F* arguments could produce unintended conclusions such as the 'Powerful God'-explanation. We have presupposed the existence of such a background theory. The question how to discover and corroborate such general theories is beyond the scope of this article.

Law-Finding-From-Facts is also different from abductive reasoning. 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 *L3F* arguments it is the causal law that is inferred and not the fact $P(a)$.

3 *Shoham's Non-Monotonic Epistemic Logic (NMEL)*

In this section we will give the syntax and semantics of Shoham's non-monotonic epistemic logical *NMEL* as it was defined in [Sho88]. *NMEL* is a point-based temporal logic augmented by the modal operators $\Box$ and $\Diamond$.

NMEL is a slightly simplified version of Shoham's non-monotonic epistemic logic *CI* as it is defined in [Sho88a, b]. *NMEL* is simpler than *CI*, because *NMEL* is point-based, whereas Shoham's logic is interval-based. This simplification is not essential for this paper.

Syntax of NMEL

Let P be a set of primitive propositions, TV a set of temporal variables, e.g. t, TC the set of temporal constants $\{..., -2, -1, 0, 1, 2, ...\} \cup \{t'\}$, U the union $TC \cup TV$, 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 and ψ are wffs, then so are $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$, $\varphi \rightarrow \psi$, $\Box\varphi$ and $\Diamond\varphi$.
4) If φ is a wff and $t \in TV$, then $\forall t\varphi$ and $\exists t\varphi$ are also wff.

We say that an *NMEL*-formula is *base* if it does not contain the modal operators $\Box$ or $\Diamond$. With respect to negation we have the following convention $(u, \neg\varphi) \leftrightarrow \neg(u, \varphi)$. In expressions like $(1, S_1 = \text{on})$ the subexpression $S_1 = \text{on}$ is considered to be a primitive proposition, 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. *NMEL* 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:

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

Readers familiar with Hintikka's epistemic logic [Hin62] will observe that $\Diamond$ is not the same as Hintikka's belief operator B. 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*, whereas the formula $B\ \varphi \wedge B\ \neg \varphi$ is inconsistent in Hintikka's epistemic logic.

Semantics of NMEL

We start with the following definitions. The symbol N is used to denote the natural numbers with the standard ordering $\leq$. A model *M* is an *S5* Kripke structure $< W, R, I >$, where *W* is a non-empty universe of possible worlds, and *I* is an interpretation function with $I\colon P \rightarrow 2^{W \times N}$. The accessibility relation *R* between worlds is universal, i.e. $\forall w, w' \in W\colon wRw'$. An example of such an *NMEL*-model is the following structure:

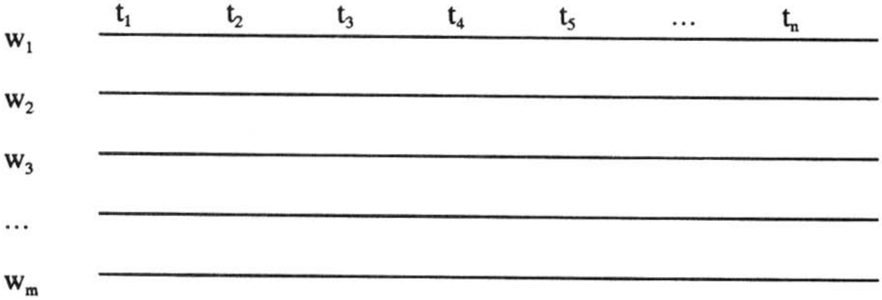

Fig. 4 Example of an *NMEL*-model

Here w_i is a possible world, and t_i a timepoint. A variable assignment is a function $VA\colon TV \rightarrow$ 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 a world w of a model M under the variable assignment VA, written $M, w \models \varphi[VA]$, under the following truth definition:

$M, w \models u_1 = u_2\ [VA]$ iff $VAL(u_1) = VAL(u_2)$
$M, w \models u_1 \leq u_2\ [VA]$ iff $VAL(u_1) \leq VAL(u_2)$
$M, w \models (u, p)\ [VA]$ iff $(w, VAL(u)) \in \mathrm{I}(p)$
$M, w \models \varphi \wedge \psi\ [VA]$ iff $M, w \models \varphi\ [VA]$ and $M, w \models \psi\ [VA]$
$M, w \models \neg\varphi\ [VA]$ iff $M, w \not\models \varphi\ [VA]$
$M, w \models \forall t\ \varphi\ [VA]$ iff $M, w \models \varphi\ [VA']$ for all alternative assignments VA' that agree with VA everywhere except possibly on t.

$M, w \models \Box \varphi\ [VA]$ iff $M, w' \models \varphi\ [VA]$ for all $w' \in W$
$M, w \models \Diamond \varphi\ [VA]$ iff $M, w' \models \varphi\ [VA]$ for at least one $w' \in W$

Because of the universal accesibilty relation we will be able to omit the specific world, and write simply $M \models \Box\varphi\ [VA]$ and $M \not\models \Box\varphi\ [VA]$ without fear of ambiguity, and analogous for $\Diamond\varphi$. Note that from the identity of time across worlds follows the validity of the so-called *Barcan formula*, i.e. $\Box\forall t\ \varphi \leftrightarrow \forall t\ \Box\varphi$. An *NMEL*-formula φ is *valid* if for all models *M* and all worlds *w* in *M* and all assignments *VA* holds $M, w \models \varphi\ [VA]$.

With respect to the 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*, also referred to as *EL*. This logic corresponds to *TK* in [Sho88a, b]. The *non-monotonic entailment* relation yields *NMEL* itself. This logic corresponds to *CI* in [Sho88a,b].

Definition 1 (*Monotonic Entailment*)

Let Φ be a set of *NMEL*-formulas, and ψ an *NMEL*-formula, Φ *monotonically entails* ψ, written $\Phi \models \psi$, if for all models *M* and all worlds *w* in *M*:

If $M, w \models \varphi\ [VA]$ for all $\varphi \in \Phi$, then $M, w \models \psi\ [VA]$.

Before we can define the non-monotonic entailment relation we first have to introduce some extra notions.

Definition 2

The *latest time point* (*ltp*) of a base formula is the latest time point 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 3
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 $M_2 \models \Box\varphi$ then also $M_1 \models \Box\varphi$, and
2) there exists some base sentence φ whose *ltp* is t such that $M_1 \models \Box\varphi$, but $M_2 \not\models \Box\varphi$

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.

Definition 4

M is a *minimal* knowledge model of φ, written $M \models_{\subset} \varphi$, if $M \models \varphi$ and there is no other M' such that $M' \models \varphi$ and $M \subset M'$.

The non-monotonic semantic entailment can now be defined as follows.

Definition 5 (Non-monotonic Entailment)
Let Φ be a set of *NMEL*-formulas, and ψ an *NMEL*-formula, Φ *non-monotonically entails* ψ, written $\Phi \models_{\subset} \psi$, if for all models M and all worlds w in M:

If M, w $\models_{\subset} \varphi$ [VA] for all $\varphi \in \Phi$, then M, w $\models \psi$ [VA].

In other words, if Φ is true in a world w of a *minimal* model M, then ψ is true in this world w.

We will now present some results that are relevant for the arguments in this paper. Subsequently we show how some of the arguments discussed in Section 1 that were not valid in classical logic can be modelled very neatly in *NMEL*.

As every *NMEL*-model has an accessibilty relation R that is universal, all the *S5* valid formulas are valid in *NMEL*.([3]) Let α, β and γ be well-formed formulas, then the following *S5*-theorems hold in *NMEL*.

([3]) For further details about *S5* modal logic see [HC82].

Proposition 6
For all *NMEL*-models M holds

$$M \models \Diamond \forall t\, [(t, \alpha) \rightarrow (t, \gamma)] \Rightarrow M \models \Diamond \forall t\, [((t, \alpha) \wedge \Diamond(t, \beta)) \rightarrow (t, \gamma)]$$

Proof. Use the *S5* valid formulas $\varphi \rightarrow (\psi \rightarrow \varphi)$ and $(\varphi \rightarrow (\psi \rightarrow \varphi)) \leftrightarrow ((\varphi \wedge \psi) \rightarrow \varphi)$.

Proposition 7
For all *NMEL*-models M holds

$$M \models \Box\forall t\, [((t, \alpha) \wedge (t, \beta)) \rightarrow (t, \gamma)] \Rightarrow M \models \forall t[(\Box(t, \alpha) \wedge \Box(t, \beta)) \rightarrow \Box(t, \gamma)]$$

Proof. 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 $\Box(\varphi \wedge \psi) \leftrightarrow (\Box\varphi \wedge \Box\psi)$.

Proposition 8
The unique set of base sentences that are known in any minimal knowledge model of a 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 [Sho88b].[4]

To illustrate the difference between *NMEL* and its monotonic version *EL*, consider the following argument. In the sequel a dotted line will indicate that the argument is valid in *NMEL*.

V. 1) B
2) $B \wedge \Diamond C \rightarrow D$
....
D

[4] Actually Shoham claims that proposition 8 only holds with respect to special theories; what he calls 'causal' theories. However, closer inspection of his proof of corollary 4.5 shows that this restriction is not necessary. Anyway, all the examples in this paper are about causal theories in Shoham's sense.

This argument is not valid in *EL*, because the premise $\Diamond\ C$ is lacking. In *NMEL* however this argument is valid. This can be argued as follows. Consider an arbitrary minimal model M of this set of premises, i.e. $M \models_c B \wedge (B \wedge \Diamond\ C) \rightarrow D$. This implies that

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

Since the formula $\Box\ \neg C$ 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. $M \not\models \Box\ \neg C$.[5] Consequently, we have $M \models \neg\Box\ \neg C$. So, by definition, it follows that

$$M \models \Diamond\ C \tag{2}$$

From (1) and (2) it follows that $M \models D$. Hence, the argument above is valid in *NMEL*.

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.

4 *Applying NMEL to Explanations Based on Incomplete Information*

After the brief introduction to *NMEL* in the previous section, we can now explain how *NMEL* can be used to formalize the explanation of person P of the burning of the lamp in the electric circuit. The problem was that although P is unable to predict the burning of the lamp, he can very well explain afterwards 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 U, 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

[5] It is assumed that B is not logically equivalent with $\Box\neg C$.

predict the burning of the lamp.

At timepoint t_2 person 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. The *L3F* argument I can be formalized in *NMEL* as follows:

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

....

$$\Diamond\ \forall t\ [((t, S_1 = \text{on}) \land \Diamond\ (t, \neg \text{DF})) \rightarrow (t, L = \text{on})] \qquad (U)$$

So, this argument formalizes the Law-Finding-From-Facts argument.[6] It expresses that, given the factual knowledge P has about the circuit at t_2, he can *assume* that the unspecified law

$$\forall t\ [((t, S_1 = \text{on}) \land \Diamond\ (t, \neg DF)) \rightarrow (t, L = \text{on})]$$

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 its monotonic version *EL*. This can be argued as follows.

Proof of argument VI:

Consider an arbitrary minimal model M which makes the premises 1 and 2 true of argument VI. Due to the minimality of M it is obvious that 1 and 2 are the only atomic base formulas known in M. It is also obvious that the formula $\Box\ \neg\forall t\ [(t, S_1 = \text{on}) \rightarrow (t, L = \text{on})]$ is not *S5*-entailed by the premises 1 and 2. Hence, due to the minimality of M and Proposition 8 it follows that

[6] The proof of VI 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 VI we mention U instead of this stronger conclusion to emphasize the analogy the *L3F* arguments in the previous section.

$M \models \neg\Box\neg\forall t\, [(t, S_1 = \text{on}) \rightarrow (t, L = \text{on})]$
$\Leftrightarrow\ M \models \Diamond\forall t\, [(t, S_1 = \text{on}) \rightarrow (t, L = \text{on})]$
Hence, by Proposition 6,
$M \models \Diamond\forall t\, [((t, S_1 = \text{on}) \wedge \Diamond(t, \neg DF)) \rightarrow (t, L = \text{on})]$

Hence the argument is valid in *NMEL*.

Argument VI is not valid in the monotonic version *EL*. This can be shown by the following counterexample. Consider a model *M*' that satisfies the formulas $\Box(2, S_1 = \text{on})$, $\Box(2, L = \text{on})$, $\Box(3, S_1 = \text{on})$ and $\Box(3, L = \text{off})$. 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. □

The *NMEL* analysis also accounts for the fact that person *P* cannot beforehand 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.

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

At t_1 person *P* may use the following *L3F* argument to find the unspecified law *U*:

VII. 1) $\Box\ (2, S_1 = \text{on})$
....
$\Diamond\forall t\, [((t, S_1 = \text{on}) \wedge \Diamond\ (t, \neg DF)) \rightarrow (t, L = \text{on})]$ (*U*)

However, the law *U* is not strong enough to derive in *NMEL* from the premise 1 that at t_2 person *P* knows that the light is on, i.e. $\Box\ (2, L = \text{on})$. Actually, the following law is needed to derive this conclusion in *NMEL* from premise 1.

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

In other words, it is not enough that the uspecified law is simply assumed, but this unspecified law itself must be really *known*. But *L3F* arguments only yield *assumed* laws, and not *known* laws. For knowledge one needs more information. A subtle but albeit crucial difference! If *P* would know the unspecified law, then we would have the following argument in *NMEL*.

VIII. 1) $\Box\ (2, S_1 = \text{on})$
2) $\Box\ \forall t\ [((t, S_1 = \text{on}) \land \Diamond\ (t, \neg DF)) \rightarrow (t, L = \text{on})]$
....
$\Box\ (2, L = \text{on})]$

This argument is valid in *NMEL* (the proof is analogous to the proof of argument V), and not in *EL*. But *P* does not know this unspecfied law, therefore, 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.

If person P had *complete knowledge* about the circuit, then he really could predict that the light will go on. This can be argued 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.

$$\Box\ \forall t\ [((t, S_1 = \text{on}) \land (t, S_2 = \text{on})) \rightarrow (t, L = \text{on})],$$

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:

IX. 1) $\Box(2, S_1 = \text{on})$
2) $\Box\ (2, S_2 = \text{on})$
3) $\Box\ \forall t\ [((t, S_1 = \text{on}) \land (t, S_2 = \text{on})) \rightarrow (t, L = \text{on})]$
....
$\Box\ (2, L = \text{on})$

Due to Proposition 7, premise 3 implies

$$\forall t\ [(\Box\ (t, S_1 = \text{on}) \land \Box\ (t, S_2 = \text{on})) \rightarrow \Box\ (t, L = \text{on})].$$

Hence, it is obvious that this predictive argument is valid in *NMEL* as well as *EL*. 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 VI, we have to make one final comment. The following explanatory argument is also valid in *NMEL*:

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

....

$\Diamond\ \forall t\ [((t, L = \text{on}) \wedge \Diamond(t, \neg DF)) \rightarrow (t, S_1 = \text{on})]$ (*CU*)

Proof of argument X:
Analogous to the proof of VI. 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. □

The difference between arguments VI and X is that in the conclusions of these arguments the causal direction is reversed; *CU* says that S_1 being on is caused by the light going on. The arguments VI and X model exacly the arguments III. 1 and III. 2 respectively, that were discussed in the Section 2. In that section we argued that the unintended unspecific law *CU* was excluded by the background theory *T*. Presupposing that *T* excludes *CU*, i.e. $T \models \neg CU$, and that *T* does not exclude *U*, i.e. $T \not\models \neg U$, we would get the following arguments in *NMEL*, which model exactly the arguments IV. 1 and IV. 2 that we discussed in the Section 2.

XI.1. 1) *T*
2) $\Box$ (2, S_1 = on)
3) $\Box$ (2, L = on)
....
U

XI.2. 1) *T*
2) $\Box$ (2, S_1 = on)
3) $\Box$ (2, L = on)
....
$\neg CU$

Proof of arguments XI.1 and XI.2:

Consider an arbitrary minimal model *M* that satisfies the premises 1, 2 and *T*. It was presupposed that $\neg CU$ is *S5*-entailed by 1, 2 and *T*, i.e. 1, 2, $T \models \neg CU$. Furthermore it is presupposed that 1, 2 and *T* do not exclude *U*, i.e. 1, 2, $T \not\models \neg U$. Hence, $\neg U$ is not *S5*-entailed by 1, 2 and *T*. If we can prove that $M \models U$ and $M \models \neg CU$, then it follows immediately that $M \models U \wedge \neg CU$. $M \models \neg CU$ holds, because $M \models 1 \wedge 2 \wedge T$ and 1, 2, $T \models \neg CU$. The proof of $M \models U$ is more complicated. First observe that 1, 2, $T \not\models \neg U$ implies that

$$1, 2, T \not\models \Box\ \neg\forall t[((t, S_1 = \text{on}) \wedge \Diamond(t, \neg DF)) \rightarrow (t, L = \text{on})].$$

Hence, there is a model N such that $N \models {}_1 \wedge 2 \wedge T$ and

$$N \not\models \Box\ \neg\forall t[((t, S_1 = \text{on}) \wedge \Diamond(t, \neg DF)) \rightarrow (t, L = \text{on})].$$

It is simple to see that then also $N \not\models \Box\ \neg\forall t[(t, S_1 = \text{on}) \rightarrow (t, L = \text{on})]$. Consequently, we also have $1, 2, T \not\models \Box\ \neg\forall t[(t, S_1 = \text{on}) \rightarrow (t, L = \text{on})]$. As this last formula is of the form $\Box\ \varphi$, with φ a base formula, we can apply Proposition 8 with the result that $M \models \neg\Box\ \neg\forall t[(t, S_1 = \text{on}) \rightarrow (t, L = \text{on})]$. This implies, due to Proposition 6 and the definition of $\Diamond$, that

$$M \models \Diamond\forall t[((t, S_1 = \text{on}) \wedge \Diamond(t, \neg DF)) \rightarrow (t, L = \text{on})].$$

Hence, we have $M \models U$. □

If we would really take all of *P*'s general background knowledge into account, we arrive at exactly the intended explanatory argument. Hence, we showed that the *L3F* arguments I, III.1, III.2, IV.1 and IV.2 that are characteristic for explanations based on incomplete information can be very neatly modelled in the non-monotonic epistemic logic *NMEL*.

4. *Conclusions*

In this paper we have shown that explanations based on incomplete information do not always comply with Hempel's covering law model. Using the example of a 'hidden' electric circuit we have shown that the relevant covering law is usually not known beforehand in this type of explanations. This results in a breakdown of the symmetry between prediction and explanation in Hempel's *DN*-model in the case of explanations based on incomplete information. We argued that in this type of explanations it is possible to derive from the observed facts a weaker type of law, which we called *unspecific laws*, that are strong enough for explanation, but too weak for prediction. Furthermore, we showed that this derivation of unspecific laws from observed facts presupposes a new type of arguments, which we called *Law-Finding-From-Facts* or *L3F* Arguments, which are supplementary to Hempel's covering law model. These *L3F* arguments cannot be modelled in classical logic, and should not be considered as inductive arguments either. We showed that Shoham's non-monotonic logic *NMEL* is an excellent logic to model these *L3F* arguments.

Erasmus University Rotterdam, Free University Amsterdam,

ACKNOWLEDGEMENTS

I want to thank Maarten Janssen for the many fruitful discussions we had about the topics in this paper and Peter Flach for his helpful comments.

REFERENCES

[Ben88] J. F. A. K. van Benthem, A Manual of Intensional Logic, Second edition revised and expanded, *CSLI Lecture Notes* Nr. 1, 1988.

[Eth88] D. W. Etherington, *Reasoning with Incomplete Information*, Pitman, London, 1988.

[Fla92] P. Flach, *A Model of Induction*, Technical Report, ISSN0924-7807, University of Tilburg, Tilburg, 1992.

[Fri69] M. Friedman, Money and the business cycles, in: M. Friedman, *The Optimum Quantity of Money and Other Essays*, London, MacMillan, 1969.

[Hem65] C. G. Hempel, Aspects of scientific explanation, in: C. G. Hempel, *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*, The Free Press, New York, 1965.

[Hin62] J. Hintikka, *Knowledge and Belief*, Cornell University Press, 1962.

[HC82] G. E. Hughes and M. J. Cresswell, *An Introduction to Modal Logic*, Methuen and Co. , 1982.

[JT91] M. C. W. Janssen and Y. H. Tan, Why Friedman's non-monotonic reasoning defies Hempel's Covering Law Model, *Synthese*, Vol. 86, 1991. pp. 255-284.

[JT92] M. C. W. Janssen and Y. H. Tan, Friedman's Permanent Income Hypothesis as an Example of Diagnostic Reasoning, *Economics and Philosophy,* Vol. 8, 1992, pp. 23-49.

[Ko88] K. Konolige, On the relation between default and autoepistemic logic, *Artificial Intelligence* 35, 1988, pp. 343-382.

[Kui86] T. Kuipers, Explanation by specification, *Logique et Analyse*, 116, 1986, pp. 509-521.

[LZ89] P. Langley and J. Zytkow, Datadriven approaches to empirical discovery, *Artificial Intelligence*, 13, 1989.

[LZSB87] P. Langley, J. Zytkow, H. Simon and G. Bradshaw, *Scientific Discovery: Computational Explorations of the Creative Processes*, MIT Press, Cambridge, 1987.

[Moo85] R. C. Moore, Semantical considerations on nonmonotonic logic, *Artificial Intelligence* 25, 1985, pp. 75-94.

[Sal90] W. Salmon, *Four Decades of Scientific Explanation*, University of Minnesota Press, 1990.

[Sho88a] Y. Shoham, *Reasoning about Change. Time and Causation from the Standpoint of Artificial Intelligence*, Boston, MIT Press, 1988.

[Sho88b] Y. Shoham, Chronological Ignorance, Experiments in nonmonotonic temporal reasoning, *Artificial Intelligence* 36, 1988, pp. 279-331.

[Ste83] W. Stegmüller, *Erklärung, Begründung, Kausalität*, second edition, Springer Verlag, Berlin, 1983.

[SS86] I. Sterling and E. Shapiro, *The Art of Prolog*, MIT Press, 1986.

[Tan88] Y. H. Tan, Explanations with incomplete information; a problem for Hempel's theory about causal explanations, in: W. Callebaut and P. Mostert (eds.), *Proceedings of the Maastricht Philosophy Conference 1987*, Delft, Eburon, 1988. (in Dutch)

[Tan90] Y. H. Tan, A Comparison of Inductive-statistical reasoning and default logic, ANTW, Vol. 82. 2, 1990, pp. 117-140. (in Dutch)