The Frame Problem: Artificial Intelligence Meets David Hume

JAMES H. FETZER
University of Minnesota, Duluth

ABSTRACT: The frame problem is a special case of the problem of induction. It has at least three aspects: (a) When can we know when something is going to change? (b) When can we know when something is not going to change? (c) What can we do when we don't know the answer to (a) or (b)? The solution to the frame problem, like the solution to the problem of induction, requires adopting an account of the nature of laws of which David Hume would have disapproved. Its solution is not merely a matter of implementation.

1. INTRODUCTION

The conundrum that is known as "the frame problem" within AI is but a special case of a very familiar problem within the theory of knowledge, a problem first emphasized by David Hume. Hume thought our knowledge of the future represented mere "habits of mind" fashioned on the basis of past experience, which create psychological expectations that are ultimately not amenable to rational warrant. The problem of induction as Hume viewed it was one of justifying some inferences about the future as opposed to others. The frame problem, likewise, is one of justifying some inferences about the future as opposed to others. The second problem is an instance of the first.

Bertrand Russell was a 20th century student of Hume's 18th century problem. Russell observed that this problem cannot be solved simply by postulating that the future will be like the past. Any postulate which asserts that the future will be like the past in *every* respect, after all, appears to be false. Indeed, change appears to be an inevitable feature of the passage of time. Any postulate which asserts that the future will be like the past in *some* respects, by comparison, appears to be true. But

International Journal of Expert Systems Pp. 219–232, Volume 3, Number 3, 1990, ISSN 0894-9077.

the problem can only be resolved if we *know* the ways in which it will change and the ways it will not. The postulate that is true is trivial, alas, while the postulate that is significant is false.

The purpose of this paper is to suggest that the frame problem, like the problem of induction before it, can be resolved by means of a theory about the nature of natural laws. This theory implies that laws cannot be violated and cannot be changed. It satisfies Russell's concern because it specifies the respects in which the future will be like the past. It agrees with Hume to the extent to which our *knowledge* about the laws of nature must always be uncertain as a product of fallible inductive reasoning. But it differs from Hume in rejecting the position that every justifiable idea has to be reducible to impressions from experience or deductive consequences that follow from them.

2. THERE ARE AT LEAST THREE ASPECTS TO "THE FRAME PROBLEM"

There are various versions of "the frame problem." Some authors even regard it as a matter of implementation in an appropriate programming language. Although issues of implementation are not irrelevant to the solution of this problem as a practical difficulty, they cannot be dealt with in the absence of resolution of the deeper problem of the extent to which the future will resemble the past. Without knowledge about the future (even though it must be fallible and uncertain), it would be impossible to resolve matters of implementation. Without a resolution to the problem of induction as it is encountered in AI, in other words, there would be no solution to implement.

Other authors have appreciated the character of this problem, whether or not they have understood its relationship to the illustrious predecessor to which it stands as a special case. Eugene Charniak and Drew McDermott, for example, have provided an appropriate depiction of its general dimensions:

> The need to infer explicitly that a state will not change across time is called the frame problem. It is a problem because almost all states fail to change during an event, and in practical systems there will be an enormous number of them, which it is impractical to deal with explicitly. This large set forms a "frame" within which a small number of changes occur, hence the phrase. (Charniak & McDermott, 1985, p. 418)

The basic problem is one of ascertaining which states change and which do not change during a temporal sequence. But that is not the only issue, even when problems of implementation are temporarily left to one side. We also need to know what to do if we don't know the answer to the basic problem.

From this perspective, what is known as "the frame problem" actually possesses at least three different aspects, which can be indicated as follows:

(a) How can we know when something is going to change?
(b) How can we know when something is not going to change?
(c) What can we do when we don't know enough to know (a) or (b)?

The answers to these questions depend upon causation and laws. Causes, of course, "bring about" changes from state $S1$ at time $t1$ to state $S2$ at time $t2$. Causal laws themselves, in turn, may be deterministic ($= u \Rightarrow$) or probabilistic ($= p \Rightarrow$), where the difference between them concerns the strength of the causal tendency for the conditions whose simultaneous presence constitutes the state $S1$ to bring about an outcome of the kind whose presence constitutes $S2$.

Whether a law is deterministic (by characterizing conditions that invariably and without exception bring a certain outcome about) or is probabilistic (by characterizing conditions that probably and only with exceptions bring a certain outcome about), the antecedents of those laws have to be maximally specific. This means that, for a specific outcome A, every factor $F1, F2, \ldots, Fm$, whose presence or absence makes a difference to the occurrence of that outcome has to be taken into account in formulating any sentence describing such a law. If even one property that makes a difference is not reflected by a corresponding predicate, a sentence that describes that law cannot be true.

The requirement of maximal specificity can be formalized and this theory of laws can be elaborated in detail (Fetzer, 1981, Pt. I). The purpose of this essay might be better served, however, by illustrating its application to a few specific cases instead. These cases exemplify the kinds of difficulties that can be encountered when the requirement of maximal specificity is overlooked or ignored. Cases of the first kind display the problem of missing conditions, of the second kind the problem of interfering (or of "counteracting") conditions, and of the third kind the problem of unusual conditions.

(1) Consider, for example, an attempt to light a match. It might be the case that the match is dry, is of a correct chemical composition, has not been exposed to rain or otherwise been made wet, etc. Nevertheless, that match would not light, no matter how many times you were to strike it, unless oxygen were present. Thus, when even one factor whose presence or absence makes a difference to the occurrence of that phenomenon is not present, the outcomes that otherwise should be expected to occur (either invariably, if the law is deterministic; or probably, if it is probabilistic) may not occur at all.

(2) Consider, for example, an attempted suicide. There are two kinds of poison, which are alkaline and acidic, respectively. Suppose that, under ordinary conditions, rather large doses of either poison would be sufficient to induce death. Consuming that amount of either poison would ordinarily bring death about. But a very serious person who wanted to guarantee his death might decide to drink large doses of them both. Since alkaline poison counteracts the effects of acidic poison, especially when consumed in equal doses, the result of this deliberate attempt to end life would not do so at all.

(3) Consider, for example, an attempted murder. Under ordinary circumstances, no doubt, stabbing a person through the heart with a knife is sufficient to bring about his death. There are, however, some unusual conditions under which that would not occur. Thus, were the victim a patient about to undergo a heart transplant operation whose blood was being circulated artificially by means of a heart-lung machine, then even though a knife were stabbed directly through his heart, that would not bring about his death at all. Such cases might be rare, but they are obviously possible.

I would like to believe that these illustrations suggest that "common sense" (or "ordinary") knowledge is not enough to provide a firm foundation for resolving the frame problem. The idea of counting on the haphazard and unsystematic accumulation of beliefs that occurs throughout our daily lives for sufficient information to anticipate the occurrence of future events is faintly ridiculous. What is required is the kind of knowledge we only possess when we have discovered those causal laws whose operation governs the processes and outcomes concerning us. There is no other way.

No problems can be solved unless we are willing to embrace answers that have the potential to solve them. As in the case of the problem of induction itself, the frame problem can only be overcome by adopting a theory of the nature of natural laws that goes beyond what Hume's theory of knowledge would permit. I have already developed the case against Hume (Fetzer, 1981, Ch. 7), and I have discussed the peculiar limitations of common sense in another place (Fetzer, 1990, Pt. II). Here I shall simply assume the positions that I have elaborated there and consider their implications.

3. HOW CAN WE KNOW WHEN SOMETHING IS GOING TO CHANGE?

The relation between causation and change, in principle, is simple and straightforward, since causes bring about changes. In practice, of course, things are more complicated, because knowledge of causal relations can be ascertained only on the basis of systematic investigation. While the aim of scientific inquiry is the discovery of natural laws, the knowledge that such research can provide is always fallible and uncertain. It is always possible—and sometimes happens—that beliefs about laws of nature are affected by later observations and experiments. While laws of nature themselves cannot change, our beliefs about what are the laws of nature may change. Nevertheless, at least two different kinds of inferential situations can arise as a function of the kind of law that subsumes the phenomena under consideration. The first presupposes that we possess knowledge of a deterministic law of the form,

$$\text{(DL-1)} \qquad (x)(t)[(F1xt \,\&\, F2xt \,\&\, \ldots \,\&\, Fmxt) = u \Rightarrow Axt^*]$$

which asserts that an occurrence of conditions $F1$, $F2$, ..., Fm (invariably, with universal strength) brings about an occurrence of an outcome of kind A (where time t^* occurs some definite interval of time later than time t).

When knowledge of a deterministic law of form (DL-1) is available, it is possible to formulate a predictive argument with a demonstrative form, which exemplifies deterministic-deductive inferential situations generally:

$$\text{(DL-2)} \qquad \begin{array}{l} (x)(t)[(F1xt \,\&\, F2xt \,\&\, \ldots \,\&\, Fmxt) = u \Rightarrow Axt^*] \\ F1at1 \,\&\, F2at1 \,\&\, \ldots \,\&\, Fmat1 \\ \hline Aat1^* \end{array} [u]$$

In this case, knowledge of the occurrence of an instance of the conditions that are specified by that law permits a deductive inference to be drawn concerning the occurrence of an outcome of a corresponding kind. The logical relationship between the premises of this prediction and its conclusion are those of complete entailment. Such arguments are deductively valid.

The second kind of inferential situation that can occur arises when laws are known, but they are not deterministic laws of kind (DL-1). These cases presuppose that we possess knowledge of a probabilistic law of the form,

$$\text{(PL-1)} \qquad (x)(t)[(F1xt \,\&\, F2xt \,\&\, \ldots \,\&\, Fmxt) = p \Rightarrow Axt^*]$$

which asserts that an occurrence of conditions $F1$, $F2$, ..., Fm (probably, with probabilistic strength) brings about an occurrence of an outcome of kind A. Thus, laws of form (PL-1), unlike those of form (DL-1), are compatible with the occurrence or the non-occurrence of outcomes of kind A.

When knowledge of a probabilistic law of form (PL-1) is available, it is possible to formulate a predictive argument of non-demonstrative form which exemplifies probabilistic-inductive inferential situations generally:

$$\text{(PL-2)} \qquad \begin{array}{l} (x)(t)[(F1xt \,\&\, F2xt \,\&\, \ldots \,\&\, Fmxt) = p \Rightarrow Axt^*] \\ F1at1 \,\&\, F2at1 \,\&\, \ldots \,\&\, Fmat1 \\ \hline\hline \\ Aat1^* \end{array} \quad [p]$$

In this case, knowledge of the occurrence of an instance of the conditions that are specified by that law permits an inductive inference to be drawn concerning the occurrence of an outcome of a corresponding kind. The logical relationship between the premises of this prediction and its conclusion are those of partial entailment. Thus, such arguments are deductively invalid.

At least two significant features of arguments of form (DL-2) and form (PL-2) deserve to be emphasized. The first is that a single line between the premises and the conclusion of (DL-2) indicates that this relationship is one of deductive validity, while the double line between the premises and the conclusion of (PL-2) indicates that this relationship is one of inductive propriety instead. The second is that the number in brackets $[n]$ indicates the degree of nomic expectability with which the truth of the conclusion ought to be expected given the truth of the premises. It reflects a nomic expectability for each single trial and approximates a relative frequency over long runs of trials.

Strictly speaking, the satisfaction of the requirement of maximal specificity means that the system thereby described is a "closed system" in relation to the occurrence of a corresponding outcome. For cases of closed systems, it is possible to predict—with deductive certainty or with probabilistic confidence—precisely how those systems will behave over the interval of time $t - t^*$ (when those properties are instantiated at time t and the outcome occurs at t^*), so long as the laws of systems of those kinds are known. Issues of implementation only become important problems at this juncture.

4. HOW CAN "CONVERSATIONAL SCOREKEEPING" HELP?

Conversational scorekeeping is a phrase describing the tacit assumptions at work in a conversation between two persons on a subject where inferences based upon premises are involved (Lewis, 1973). Ordinary conversations, of course, may transpire over an extended period of time, where various assumptions between its participants can remain constant across time. The following pattern might occur as part of a conversation:

(OC) *an ordinary conversation:*

Agree on premise (a) at $t1$: (a) p or q
Agree on premise (b) at $t2$: (b) $\neg p$
Entitled to deduce (c) at $t3$: (c) q

where $t2$ might occur long after $t1$, for example, assuming that nothing intervened in the meanwhile to affect these participants in their beliefs.

Analogous patterns of conversation can be discerned within the context of scientific discourse. The following pattern might occur as part of a scientific experiment involving the application of some established law:

(SC) *a scientific conversation:*

Assume law (L1) at $t1$:	(L1)	$(F1at1 \,\&\, \ldots \,\&\, Fmat1) = n \Rightarrow Aat1^*$
Agree about (F1) at $t2$:	(F1)	$F1at1$
Agree about (F2) at $t3$:	(F2)	$F2at1$
Agree about (...) at t ...:	(...)	...
Agree about (Fm) at tm:	(Fm)	$Fmat1$
Entitled to infer (A) at tn:	(A)	$Aat1^*$

The strength of the inferential relationship between the premises (L1)-(Fm) and the conclusion (A) in such cases depends upon and varies as a function of the strength of the causal tendency n, which in turn generates the degree of nomic expectability $[n]$ with which that conclusion follows. In deterministic-deductive inferential situations, of course, n will equal u, and in probabilistic-inductive inferential situations, n will equal p instead.

The combination of conversational scorekeeping and the requirement of maximal specificity suggests a solution for the problem of implementation. The function **cond** in LISP (Wilensky, 1984, p. 55), for example, offers one among many means for representing causal reasoning when maximally specific causal antecedents are known. The **cond** function, (Fun 1),

```
(Fun1)  (cond  (exp 11 exp 12 exp 13 ...)
               (exp 21 exp 22 exp 23 ...)
               (exp 31 exp 32 exp 33 ...)
               ⋮
               (exp n1 exp n2 exp n3 ...))
```

executes in the following fashion. LISP examines the first **cond** clause, where each sequence (**exp 11 exp 12 exp 13** ...) qualifies as one clause. If the first element of that clause is true (not-**nil**), it continues down that clause until it comes to the end. When they are all true, LISP returns the last element of that clause as its value and, otherwise, **nil**.

This function can be readily adapted for the purpose of implementing the patterns of reasoning that we have been discussing here. Notice, in particular, that it can implement patterns of conversation such as (SC):

```
(Fun2)  (cond  (F1xt  F2xt  ...  Fmxt  Axt*)
               (F1xt  -F2xt  ...  -Fmxt  -Axt*)
               (-F1xt  F2xt  ...  Fmxt  Axt*)
               ⋮
               (-F1xt -F2xt ... -Fmxt -Axt*))
```

The use of functions such as **cond**, therefore, can be employed to implement representations of scientific knowledge of causal laws of the kind (DL-1) and, with suitable enhancements, of the kind (PL-1). This illustration implies that even issues of implementation can likewise be resolved.

5. HOW CAN WE KNOW WHEN SOMETHING IS NOT GOING TO CHANGE?

The second version of "the frame problem," by comparison, concerns when we can know that something is *not* going to change. The persistence of a specific factor *Fi* from time *t* to time *t** itself can be subjected to systematic prediction when we possess knowledge of all those factors whose presence or absence would make a difference to the presence or absence of *Fi* during that interval. Suppose, for example, that a deterministic law is known which relates the existence of *Fi* at *t* to the existence of *Fi* at *t**:

$$\text{(DL-3)} \qquad (x)(t)[(Fixt \;\&\; Txt) = u \Rightarrow Fixt^*]$$

A law of this form asserts that conditions *Fi* continue to endure through the temporal interval from *t* to *t**, so long as condition *T* also obtains at *t*.

When knowledge of a deterministic law of form (DL-3) is available, it is possible to formulate a predictive argument with a demonstrative form that predicts that *Fi* instantiated at time *t* will remained unchanged at *t**:

$$\text{(DL-4)} \qquad \begin{array}{l} (x)(t)[(F1xt \,\&\, Txt) = u \Rightarrow Fixt^*] \\ Fiat1 \,\&\, Tat1 \\ \hline Fiat1^* \end{array} [u]$$

where "*T*" specifies the presence or the absence of each factor whose presence or absence makes a difference to the possibility that *Fi* might change or not (i.e., every relevant property in relation to that outcome from *t* to *t**).

As a simple example, note that an instance of an argument of the deductive form (DL-4) might predict that the barbecue coals would continue to burn from the time they were ignited at *t*1 to the time the steaks were placed on the grill at *t*2, provided that they were burning properly at *t*1 and "nothing happened" (say, a rain storm, an errant hosing-down, etc.) to bring that process to an end. But laws of this kind, like all other scientific laws, must be maximally specific and overlook no relevant property at all. There need to have been enough properly-ignited coals to begin with, etc.

Once again, there are probabilistic counterparts. A probabilistic law might be known relating the existence of *Fi* at *t* to the existence of *Fi* at *t**:

$$\text{(PL-3)} \qquad (x)(t)[(Fixt \,\&\, Txt) = p \Rightarrow Fixt^*]$$

A law of this form asserts that conditions *Fi* tend to endure (with probability *p*) through the temporal interval from *t* to *t**, provided condition *T* obtains at *t* also. The truth of a probabilistic law of form (PL-3) not only provides no guarantee that the conditions *Fi* must continue to endure but even implies that they will not continue to endure with probability $1 - p$!

When knowledge of a probabilistic law of form (PL-3) is available, it is possible to formulate a predictive argument with a non-demonstrative form predicting that *Fi* instantiated at *t* will probably be unchanged at *t**:

$$\text{(PL-4)} \qquad \begin{array}{l} (x)(t)[(Fixt \,\&\, Txt) = p \Rightarrow Fixt^*] \\ Fiat1 \,\&\, Tat1 \\ \hline\hline Fiat1^* \end{array} [p]$$

where, as before, "*T*" specifies the presence or the absence of each factor whose presence or absence makes a difference to the possibility that *Fi* might change. Corresponding arguments could be constructed for alternative outcomes that might occur in lieu of *Fi* under the same conditions.

As a less simple example, note that an instance of an argument of the inductive form (PL-4) might predict that a particular atom of polonium^{218} at time t would remain intact by time t^*, where $t^* = t + 3.05$ minutes (with probability p equal to one-half), provided that it is not subjected to nuclear bombardment, etc., because the half-life of polonium^{218} is 3.05 minutes. A corresponding argument of a similar form might predict that that same atom would undergo decay during that same temporal interval (with probability one minus p), under the very same conditions, of course, because this is a probabilistic, rather than a deterministic, phenomenon.

Notice that the laws and arguments considered in Section 3 were all concerned with closed systems, relative to which a specific outcome event, A, will or will not occur. The laws and arguments considered in Section 4 are also concerned with closed systems, relative to which a specific set of conditions, such as *Fi*, may or may not endure from time t to time t^*. Indeed, even when *Fiat* happens to be a closed system in relation to an outcome of kind *Aat**, this does not mean that those conditions must endure forever. A sugar cube left in its box on the shelf in a dry pantry, for example, tends to persist as a cube of sugar, but not when it's dropped into a cup of hot coffee.

Perhaps it ought to be emphasized that these argument forms serve to establish the degree of nomic expectability with which the occurrence described by their conclusions should be expected, given the truth of the sentences that constitute their premises. They do not indicate whether a degree of nomic expectability of .6, for example, is strong enough to support a specific decision. When the prospects for fair weather are .6, the prospects for other-than-fair weather are .4. Whether or not the family should depart for a picnic under these conditions concerns decision-making policies that go far beyond the scope of nomic expectabilities as such.

6. WHAT CAN WE DO WHEN WE DON'T KNOW ENOUGH TO KNOW (A) OR (B)?

The inferential situation is somewhat more complicated than it has been described thus far, because predictions can be made on the basis of statistical knowledge of relative frequencies in the past as well as on the basis of inductive knowledge of natural laws. The benefit of inferences drawn on the basis of natural laws, however, is that laws as properties of the world cannot be changed and cannot be violated. Laws that have obtained in the past will obtain in the future, necessarily, as functions of the "permanent property" relations which they embody (Fetzer, 1981).

The hazard with inferences drawn on the basis of relative frequencies, therefore, is that relative frequencies as properties of the world can be violated and can be changed. Frequencies that have obtained in the past need not obtain in the future, necessarily, as functions of the "transient property" relations that they may represent. That 100%/50%/whatever% of the Volkswagens sold in America in the past have been painted grey, for example, provides no guarantee that 100%/50%/whatever% of the Volkswagens sold in America in the future will also be painted grey.

The underlying distinction is one between relations between properties (such as the atomic number of things that are gold and their malleability, conductivity, melting points, boiling points, etc.) that cannot be violated and cannot be changed (because there are no processes or procedures, natural or contrived, that could separate instances of the atomic number 79 and the possession of those attributes) and those that can be violated and can be changed (because there are processes or procedures, such as repainting a car), that can separate instances of those properties (such as being a Volkswagen) from those attributes (such as being painted grey).

We can refer to this situation as reflecting the *predictive primacy of scientific knowledge* (of natural laws) over other kinds of beliefs concerning the future, however rational. Nevertheless, there are conditions within which scientific knowledge (of natural laws) may be unavailable, while knowledge of relative frequencies, especially, may be available. Suppose, for example, that the causally relevant factors that bring about outcomes of a certain kind are either not finite or not known. Then the best knowledge available to us may turn out to be knowledge of relative frequencies.

Under these conditions, the strongest forms of inference that turn out to be possible may have to be based upon assumptions concerning so-called normal, typical, or standard situations. "Case-based" and "script-based" reasoning, in fact, exemplify this practice. From a statistical point of view, of course, these "norms" might be defined as means, as modes, or as medians. In any case, they are amenable to exceptions, even if they happen to reflect constant conjunctions (100% relative frequencies) that have had no exceptions in the past. Unless these relations are displays of natural laws, there are conditions under which they can be violated or can be changed.

The different kinds of inferential situations that can be encountered as a function of the different kinds of knowledge that might be available, therefore, can be summarized by means of a series of epistemic theorems. When we possess knowledge of causal laws and specific conditions, for example, the following epistemic theorems obtain in a knowledge context K:

(ET-1) When '$(x)(t)[(F1xt \ \& \ F2xt \ \& \ \ldots \ \& \ Fmxt) = u \Rightarrow Axt^*]$' and '$F1at1 \ \& \ F2at1 \ \& \ \ldots \ \& \ Fmat1$' belong to a knowledge context K, then the nomic expectability of '$Aat1^*$' in K is u;

(ET-2) When '$(x)(t)[(F1xt \ \& \ F2xt \ \& \ \ldots \ \& \ Fmxt) = p \Rightarrow Axt^*]$' and '$F1at1 \ \& \ F2at1 \ \& \ \ldots \ \& \ Fmat1$' belong to a knowledge context K, then the nomic expectability of '$Aat1^*$' in K is p; and,

(ET-3) When '$(x)(t)[(F1xt \ \& \ F2xt \ \& \ \ldots \ \& \ Fmxt) = u \Rightarrow Aat^*]$' and '$-Aat1^*$' belong to a knowledge context K, then the nomic expectability of '$-(F1at1 \ \& \ F2at1 \ \& \ \ldots \ \& \ Fmat1)$' in K is u.

When we lack knowledge of causal laws but possess knowledge of relative frequencies, by contrast, then the following epistemic theorem obtains:

(ET-4) When '$P(Axt^*/F1xt \ \& \ F2xt \ \& \ \ldots \ \& \ Fmxt) = f$' and '$F1at1 \ \& \ F2at1 \ \& \ \ldots \ \& \ Fmat1$' belong to a knowledge context K, then the "qualified-instance" expectability of '$Aat1^*$' in K is f,

where '$P(\ldots/\_\_) = f$' stands for the relative frequency f with which properties of kind ... occur in relation to properties of kind ___ and "qualified-instance" expectability reflects the best available estimate of the relative frequency for outcomes of that kind within that reference class. These values, in turn, can be utilized as "weights" for the next single case (Fetzer, 1983). (ET-4) thus affords a principle of predictive inference that can be employed on the basis of empirical knowledge of mere relations of relative frequency.

7. WHAT GENERAL CONSEQUENCES FOLLOW FROM THIS ANALYSIS?

Our knowledge of natural laws, of course, arises from empirical procedures employing inductive inference and can never be certain or be infallible. The most difficult task that science confronts, from this point of view, is that of separating good guesses from bad guesses about the laws of nature. And, as Karl Popper, especially, has emphasized, this process involves testing laws by attempting to refute them (Popper, 1968). Only by sincere attempts to show that a guess is wrong can we accumulate evidence that a guess might be right. Such evidence is always inconclusive, however, precisely because, when our guesses survive our best efforts to refute them, this does not show that they are correct. We may have not yet discovered how to refute them!

This means that a proper understanding of Popperian methodology entails appreciating the difference between the negative significance of successful attempts to refute an hypothesis, on the one hand, and the positive significance of unsuccessful attempts to refute an hypothesis, on the other. Although the tentative elimination of mistaken guesses about natural laws is an indispensable element of scientific procedure, deductive inference alone cannot possibly be enough. Without the tentative acceptance of hypotheses that have withstood our best efforts to refute them, it would be impossible to acquire the kind of knowledge that we need to explain the occurrence of events in the past or even to predict the occurrence of events in the future.

The crucial ingredient in resolving Hume's problem of induction is thus the same crucial ingredient required to solve the frame problem. Knowledge of natural laws exceeds the epistemic resources that Hume would permit, a fateful blunder that has affected the history of philosophy ever since. For Hume insisted that every justifiable idea has to be reducible to impressions from experience or *deductive* consequences that follow from them. But he should have insisted instead that every justifiable idea has to be reducible to impressions from experience or *inductive* consequences that follow from them. That would support inference to natural laws, even if our knowledge of laws will always be uncertain as a product of fallible inductive reasoning.

This essay must conclude with at least two important qualifications. An exact analysis of the problem of induction would require refinements which have been ignored within this context. These are discussed in other places, however, to which references have been made. A complete analysis of the frame problem, similarly, would acknowledge that unsolved (and possibly insoluble) issues of implementation yet remain. When laws involve more than finitely many relevant properties (if any of

them do) or only a finite but large number that exceeds the programming capacity of our available machines, then the frame problem cannot be solved in practice. Whether or not the frame problem can be solved, therefore, ultimately depends upon its specific formulation. But the solution to the frame problem is never just a matter of implementation in an appropriate programming language.

APPENDIX A: HOW ARE PREDICTIONS RELATED TO EXPLANATIONS?

The most influential account concerning the nature of arguments involving laws is known as the "covering law theory." Its principal proponent has been Carl G. Hempel, who, in a sequence of important papers, has argued for what is referred to as "the symmetry thesis" (Hempel, 1965). According to the symmetry thesis, every adequate explanation is potentially an adequate prediction (i.e., the same premises that explain the occurrence of an outcome could have served as premises for its prediction, were they taken account of in time, and conversely). In his later work, however, Hempel acknowledged that some premises suitable for prediction are not suitable for explanation.

Hempel's observation has a great deal of relevance for the views which have been presented here. Predictive inferences that are based upon relative frequencies, such as those supported by (ET-4), for example, are entirely lacking in explanatory significance. This may appear obvious, since these predictions do not involve inferences from laws. Other instances of predictions that do involve inferences from laws, such as those supported by (ET-3), however, are also lacking in explanatory significance, which is far more surprising. Indeed, cases of this kind provide conclusive evidence that the symmetry thesis itself, although plausible, ultimately cannot be sustained.

Consider, for example, every law concerning those conditions relative to which the death of a human being would be brought about. In order to emphasize the point at issue, let us restrict our attention only to sufficient conditions, such as being stepped on by an elephant (being run over by a steamroller, etc.). Every human being is such that, if he were stepped on by an elephant (were run over by a steamroller, etc.), his death would be brought about thereby. For all of us who are still alive, therefore, an inference from laws can be made to the conclusion that we have not been stepped on by an elephant (and have not been run over by a steamroller, etc.).

These arguments, of course, satisfy the conditions called for by inferences that satisfy (ET-3), which exemplifies a special case of the principle known as *modus tollens*. Arguments involving inferences that satisfy (ET-1) and (ET-2), by comparison, can fulfill the appropriate conditions for an adequate prediction to be an adequate explanation as well, so long as their premises are not just *believed* to be true but actually are true. Arguments that satisfy (ET-3), however, fulfill the conditions that are required to be an adequate prediction but not those required to be an adequate explanation, even when their premises are *true* and are not merely believed to be true.

So the symmetry thesis itself is not tenable. Even the thesis that every argument involving inferences from laws is a potential explanation turns out to be unjustifiable. If you have any doubts, ask yourself what it would take to explain *why* you have never been stepped on by an elephant (run over by a steamroller, etc.). On the other hand, arguments involving inferences from relative frequencies are *never* explanatory. That an argument involves an inference from laws turns out to be a necessary, but not a sufficient, condition for it to be an explanation. There is more to explanation than inference from laws, just as there is more to nomic explainability than nomic expectability.

APPENDIX B: WHAT DOES IT TAKE FOR AN EXPLANATION TO BE ADEQUATE?

Another important respect in which prediction differs from explanation suggests another reason why inferences from laws and nomic expectability may not produce adequate explanations. The prediction that a large sugar cube would dissolve were it dropped into a cup of hot coffee should not be faulted merely because it refers to certain properties (the size of the sugar cube, the shape of the sugar, etc.) whose presence or absence makes no difference to the occurrence of that outcome: (refined) sugar, after all, would dissolve in a cup of hot coffee, whether it were a large cube, a spoonful, etc. (Even the amount of coffee and its temperature may be causally irrelevant—at least, as long as it is not frozen!)

Although we might argue about specific features of this example—say, if the sugar cube were larger than the cup and therefore could not fit into it, thereby preventing it from dissolving, then its size would be relevant—the issue that arises here ought to be clear. In the case of *predictions*, the presence or absence of properties whose presence or absence makes no difference to the occurrence of an outcome is permissible and innocuous. But in the case of *explanations*, the presence of properties whose presence or absence makes no difference to the occurrence of their outcome is misleading and unacceptable. (See, for example, Salmon, 1971, pp. 29-87.)

The requirement of maximal specificity (RMS) itself insures that every sentence that describes a law can be true only when the presence or absence of *every* property whose presence or absence makes a difference to the occurrence of the outcome of interest is included in the antecedent of that law. The requirement of strict maximal specificity (RSMS) insures that *only* properties whose presence or absence makes a difference to the occurrence of the outcome of interest can be included in the antecedent of laws that appear in the premises of arguments that are intended to be explanatory. The requirement of strict maximal specificity is what we need.

Indeed, once this additional condition has been acknowledged, it is possible to formalize the requirements that an adequate explanation must satisfy. Because explanations are arguments which have premises and conclusions, it has become standard terminology within the theory of explanation (following Hempel) to refer to the premises of an explanation as its "explanans" and its conclusion as its "explanandum." An adequate explanation, thus understood, must satisfy the following four general requirements, namely:

(CA) A set of sentences *S*, known as "the explanans," provides an adequate (nomically significant) scientific explanation for the occurrence of a singular event described by another sentence *E*, known as "its explanandum," in relation to the language *L*, if and only if:

(CA-1) the explanandum *E* is a deductive or a probabilistic consequence of its explanans *S*;

(CA-2) the explanans *S* contains at least one lawful sentence that is actually required for the deductive or probabilistic derivation of the explanandum *E* from its explanans;

(CA-3) the explanation satisfies the requirement of strict maximal specificity with respect to its lawful premises; and,

(CA-4) the sentences constituting the explanation–both the explanans *S* and the explanandum *E*–must be true, in relation to the language *L*.

Thus, the combined force of these requirements insures that the lawful premises of an adequate explanation must specify all and only those properties whose presence or absence made a difference to the occurrence of its explanandum-phenomenon. (For discussion of these issues in greater detail, see especially Fetzer, 1981, 1987; and references mentioned there.)

Acknowledgments: The author is indebted to Paul Humphreys for several valuable suggestions, especially for hinting that two sections would function better as appendices.

REFERENCES

Charniak, E. & McDermott, D. (1985). *Introduction to Artificial Intelligence*. Reading, MA: Addison-Wesley.

Fetzer, J. H. (1981). *Scientific Knowledge*. Dordrecht, Holland: D. Reidel.

Fetzer, J. H. (1983). Probabilistic explanations. In P. Asquith & T. Nickles (Eds.), *PSA 1982*, Volume 2 (pp. 194-207). East Lansing, MI: Philosophy of Science Association.

Fetzer, J. H. (1987). Critical notice: Wesley Salmon's Scientific explanation and the causal structure of the world. *Philosophy of Science, 54*, 597-610.

Fetzer, J. H. (1990). *Artificial Intelligence: Its Scope and Limits*. Dordrecht, Holland: Kluwer.

Hempel, C. G. (1965). *Aspects of Scientific Explanation*. New York: The Free Press.

Lewis, D. (1973). Scorekeeping in a language game. *Journal of Philosophical Logic, 8*, 339-359.

Popper, K. R. (1968). *Conjectures and Refutations*. New York: Harper & Row.

Salmon, W. C. (1971). *Statistical Explanation and Statistical Relevance*. Pittsburgh, PA: University of Pittsburgh Press.

Wilensky, R. (1984). *LISPcraft*. New York: W. W. Norton.

Address all correspondence to: James H. Fetzer, Department of Philosophy, University of Minnesota, 10 University Drive, Duluth, MN 55812-2496. Email: jfetzer@ub.d.umn.edu