The Problem with Solutions to the Frame Problem

Leora Morgenstern
IBM T.J. Watson Research Center
30 Saw Mill River Road
Hawthorne, N.Y. 10532
leora@watson.ibm.com

1 Introduction

The frame problem, the problem of efficiently determining which things remain the same in a changing world, [1] has been with us for over a quarter of a century – ever since the publication of McCarthy and Hayes's famous essay, "Some Philosophical Problems from the Standpoint of Artificial Intelligence," in 1969. A quarter of a century is a very long time in the time frame of computer science, and especially in the history of Artificial Intelligence (AI), which has itself been around for only about 40 years. Indeed, it is not much younger than the advent of logicist AI (McCarthy, 1958), that brand of AI which attempts to formalize reasoning, particularly common-sense reasoning, within mathematical logic. Present since the early years of AI, the frame problem has festered within the AI community. It has drawn, and continues to draw, manpower and talent from the pool of AI researchers, particularly from the logicist community. It has pitted logicists against antilogicists, who argue that the attempt to capture common-sense reasoning within formal logic is doomed to failure. In fact, it was a major factor in the rather public and shocking (for AI standards) conversion of a stalwart logicist, Drew McDermott, to a confirmed antilogicist (McDermott, 1987).

Twenty-six years may not seem like a long time to philosophers and mathematicians

[1] This is a loose, and not quite precise, definition of the frame problem. In fact, one of the problems with the frame problem is the fact that everybody seems to define it a little differently. See section 3 for a serious investigation of what exactly the frame problem is.

(presumably the colleagues and spiritual brethren of AI logicists), whose problems are often so much older. Fermat's Last Theorem (that is, the problem of whether it is or is not a theorem) stayed with us for over three centuries before it was solved (Wiles, 1995). Philosophers have lived for over 2,000 years with the Liar Paradox. The problems that arise from the inherent inconsistency of determinism and free will have been around for at least that long. But these problems do not threaten the mathematical and philosophical communities in any way. It would seem rather ridiculous if a group of mathematicians, after spending twenty-five years on Fermat's Last Theorem, had thrown up their hands and declared that it was quite impossible to come up with a solution and that there was no point working on it anymore, or if some philosophers declared the entire endeavor of analytic philosophy to be quite worthless due to their inability to find a solution to the Liar Paradox. Yet AI researchers have publicly declared that the failure of logicist AI to produce a solution to the frame problem has amply demonstrated the inadequacy of the logicist approach.

Lest it be argued that it is precisely the venerable age of classic mathematical and philosophical problems that has protected them from the kind of furor accompanying the analysis of solutions to the frame problem, it should be noted that there are plenty of relatively young and unsolved problems that are not nearly as derided as the frame problem. The Paradoxes of Confirmation have stood as a challenge to the philosophy of science since the mid-1940s (Hempel, 1945). The $\mathbf{P} = \mathbf{NP}$? question has remained unsolved since it was first presented and conjectured that $\mathbf{P} \neq \mathbf{NP}$ (Cobham, 1964; Edmonds, 1965; Cook, 1971). Yet philosophers of science and theoretical computer scientists are not threatened, or humiliated, or torn apart by the lack of solutions.

What, then, makes the frame problem different? What is to account for the feeling of failure on the part of those AI researchers who have tried so hard to solve the frame problem? There are four reasons:

[1] *The frame problem should not be as difficult as other long-standing problems.* Many open problems are genuinely unsolvable. A paradox arises because several deeply seated intuitions turn out to be mutually inconsistent. The Liar Paradox, for example, is paradoxical precisely because some of our deeply believed intuitions about language, such as:

(i) that a language should be able to include a fully expressive quotation construct and its own truth predicate
(ii) that if a sentence is said to be true, it is in fact the case
(iii) that a sentence is either true or false
are inconsistent.
Likewise, the Paradoxes of Confirmation exist because our intuition that a black raven confirms, and a white swan does not confirm, the hypothesis that all ravens are black directly contradicts our intuition that if an object confirms one statement, it confirms all logically equivalent formulations of that statement.

So it is not surprising that years of examining the problem (in the case of the Liar Paradox, thousands of them) have not yielded a completely satisfactory solution: all solutions will to some extent seem unsatisfactory because they will violate at least one intuition. The same can be said for the dilemma of determinacy versus free will. These two concepts seem so inherently opposed that any attempt to resolve them smacks of casuistry and thus is not satisfactory.

Other open problems are just inherently difficult. Intuitions on the equivalence of deterministic and nondeterministic linear bounded automata are often weak. Even in cases in which one has a strong intuition about an open problem (it has been conjectured from the start that **P** is not equal to **NP**), there is a wide gap between the tenuousness of an intuition (often based on the *consequences* of a particular solution to an open problem) and the solid reasoning needed for a formal proof.

However, all this does *not* apply to the frame problem. The frame problem, after all, is quite straightforward, even mundane. Every five-year-old can reason that most things stay the same as time passes – and, presumably, without having explicit knowledge of, or reasoning about, all the things that do not change. The reasoning process seems uncomplicated. Thus the frustration of AI researchers who cannot capture this straightforward reasoning within formal logic is all the more acute.
[2] *The frame problem is one of the core problems of temporal reasoning.* As such, solving the frame problem is necessary for the future development of logicist AI.

Many open problems, although they may have enormous practical import, can be factored out of their disciplines. Solutions are not necessary for future progress. For example, the Liar Paradox challenges philosophers of language, since any theory based on a language that contains its own truth predicate and quotation can be inconsistent. But researchers can proceed by assuming one of the (not-so-satisfactory) suggested resolutions, and subsequently ignore the issue. Fermat's Last Theorem is certainly interesting, but it does not lie at the core of mathematics. A lot of interesting mathematics was possible before its solution.

In contrast, solving the frame problem is central to temporal reasoning, and temporal reasoning is central to AI. It is necessary for planning, explanation, and diagnosis. It arises in and is central to virtually every interesting area of AI. This is precisely why many researchers have shunned the logicist approach entirely and opt to be proceduralists. Proceduralists can ignore the frame problem, as we see in section 5.2, although this approach is sometimes not entirely correct.

[3] *So much has been written; so little progress has been made.*

The sheer bulk of published research papers on the frame problem is stupefying. There have been entire symposiums devoted to the subject, and large sections of major conferences still focus exclusively on the frame problem. Yet there is little feeling that we have made genuine progress.

Progress here is not synonymous with solution. Even when difficult problems are not solved, there can be a definite feeling of progress. Research on the Liar Paradox, for example, has made explicit the inconsistent set of intuitions underlying the paradox, suggested novel methods of (partial) resolution (Kripke, 1975), studied the consequences of various attempts at a resolution, and identified families of paradoxes (see, e.g., the collection of papers in Martin, 1984). Thus, there has been both intellectual and practical progress.

In contrast, the work on the frame problem seems Sisyphean in its nature. The general pattern of the research looks something like this: A toy problem, presumably meant to be a characteristic instance of the frame problem, is introduced; a solution is proposed; a new variant on the toy problem is introduced; it is discovered that the old solution cannot solve the new toy problem, and so on.

Now, there may be superficial similarities between this scenario and the description that

Kuhn (1970) gave of scientific research: There exists an unsolved set of problems; a solution (the first of a sequence of solutions within a paradigm) is proposed; it is discovered that the solution cannot solve some related problem(s); so the solution is modified (and is the next in the sequence in that paradigm); and so on, until this paradigm becomes useless, and a new one is suggested.

But the analogy does not hold up. The smallest twiddle on a toy problem frequently invalidates all known solutions to the original toy problem; clearly, these solutions were very narrow.

It is this lack of robustness which has frustrated logicists and disillusioned the AI community. The expression of this feeling is evident in McDermott's (1987) comment on one of the attempted solutions to the frame problem (Lifschitz, 1987), which was then believed to be an adequate solution: What happens the next time? Will Lifschitz always be there to bail us out?

Nobody has very much confidence that the attempts at solutions that are around now work on anything but a small set of toy problems. There is no sense that these "solutions" have in any way attacked the essence of the frame problem.

[4] *Computer scientists are used to getting things done.* (This does not apply to theoretical computer scientists, who are just mathematicians in disguise.) If a program does not do what its designer intended, the computer scientist will modify the program so that it does its intended task. The frame problem is all about getting things done, or more precisely, about making the proper inferences. It is the problem of determining what stays the same about the world as time passes and actions are performed – without having to explicitly state all the things that stay the same. The inability to simply code up the axioms that will allow these inferences is a source of fathomless frustration to the AI researchers who aim to solve this problem. It goes against the whole hacker experience (and even AI logicists have some of the hacker mentality in their blood).

This is *not* the experience of the philosopher whose task is often to explicate underlying issues and not necessarily to provide solutions. Mathematicians, of course, are supposed to solve problems – but they are more inured to failure. Computer scientists have not gotten this far.

1.1 Where do we go from here?

So we have established that the AI researcher is perfectly justified in feeling depressed and frustrated and disillusioned about the current state of research in the frame problem. What, then, are we to do? There are two approaches. One is to admit defeat – to resign oneself to the belief that no solution to the frame problem is possible and to proceed with research as best as one can. That is, to try to *ignore* the frame problem as much as possible. Now, it was argued earlier that one cannot ignore the frame problem; that it is too central to the problem of temporal reasoning within logicist AI to just ignore. Thus, admitting defeat for the frame problem entails admitting defeat for the entire logicist enterprise. The sole alternative is modeling temporal reasoning within proceduralist AI. Drew McDermott is one of the many who have chosen this approach.

The second approach is posited on the notion that genuine progress on the frame problem is possible. A direct consequence of this supposition is that the existing twenty-odd years of research on the frame problem [2] must be flawed in some respects. The idea, then, is to investigate what these flaws are, to determine how to avoid these flaws, and to suggest a course of research in which genuine progress is possible. This second approach is the one in which I believe, and the remainder of this chapter is dedicated to its advancement.

Why have I chosen this approach? Primarily, because of a deep commitment to, and a genuine belief in, logicist AI: the belief that common-sense reasoning can be captured within a formal language. The evidence seems strong and has been articulated by many (Hayes, 1977; McDermott, 1978, before his apostasy; Newell, 1981): A formal language like logic gives us the ability to give a *semantics*, or meaning, to the sentences we use; it gives us rules of inference that are sound and complete and mirror at least some of the reasoning that we do. It allows us to reason *about* our reasoning processes in a coherent way, in much the same way that articulate people can describe their reasoning processes within natural language. For all of its drawbacks, it is the richest, most correct language and system of reasoning that we have. To give up on this commitment, simply because nothing has worked for twenty-six years, seems like taking the easy way out. Related to

[2] Most of which has taken place since the publication of the Yale Shooting Problem (Hanks and McDermott, 1986)

this point is the suspicion that procedural AI may be easier in the short run, but is not the panacea that some assume. In particular, if a program is not solidly based upon a theory, it is likely to be ad hoc (and if it *is* based on a formal theory, we have not given up logicist AI). Finally, there is an element of sheer orneriness here – we are not going to quit just because the going gets a little tough.

The remainder of this chapter is organized as follows: First we will attempt to define what the frame problem really is. As we shall see, there is no single definition of the frame problem, since many researchers have defined the frame problem in different ways. We present a collection of such definitions, organized from the most specific to the most general. It is our contention that many of the errors in research on the frame problem arise from a misunderstanding of the level of generality on which to approach the problem. Next, we present a minimum list of requirements that every solution to the frame problem must satisfy. We then examine several of the trends in solutions to the frame problem and show where they fall short. We subsequently suggest a proper framework for research. Finally, we examine the lessons that can be learned from the frame problem to AI research in general.

2 The Origins of the Frame Problem

The term *frame problem* was originally coined to describe a technical problem that arises in a very narrow context. In the late 1950s, John McCarthy (1958) argued that common-sense reasoning could be formalized within first-order logic; in the mid-1960s he developed the *situation calculus*, an instance of first-order logic especially formulated for reasoning about time (described in McCarthy and Hayes, 1969). For our purposes, the situation calculus, a discrete branching timeline of situations connected by actions, can be summarized as follows.

We think of a situation as being a snapshot of the world at a particular instant of time. Consider, e.g., 1 p.m. EDT, July 17, 1992. In that situation, George Bush is president of the United States, Bill Clinton is the presidential nominee of the Democratic party, it is hot and muggy in New York City, and warm, dry, and sunny in Palo Alto. A *state* is defined as a collection of situations. For example, the state of Richard Nixon being president is the collection of situations starting on January 20, 1969, and ending on August 9, 1974. The state *On(BlockA, BlockB)* is the collection of all situations in which Block A is on top of Block B. A state is one type of *fluent*; intuitively, something whose value changes over time. States are Boolean fluents; other fluents are time-varying terms. For example, *President(USA)* is a fluent; it has the value George Bush in 1989, and has the value Bill Clinton starting in January, 1993. The predicates *Holds* or *True* relate fluents and situations. Thus we can say *Holds(S77,On(BlockA,BlockB))*.

Actions are defined as functions on states. For example, the action *Puton(BlockA, BlockB)* maps those states in which Block A and Block B are *clear* (with nothing on top of them) to those states in which Block A is on top of Block B. *Result* is a function mapping the situation before the action has occurred to the situation after the action has occurred. So, if we have the following statements in our theory:

Holds(S0, Clear(BlockA))
Holds(S0, Clear(BlockB))

we can talk about *Result(Puton(BlockA, BlockB), S0)*.

The situation calculus gives us a way to talk about causation. To state that the *Puton(BlockA, BlockB)* action causes Block A to be on Block B, we say that in the situation resulting from the *Puton* action, Block A is on Block B. That is:
Holds(Result(Puton(BlockA, BlockB), S0), On(BlockA, BlockB))

Thus far, everything is quite straightforward. But now, things become complicated. Suppose we start out with the following facts:
Holds(S0, Red(BlockA))
Holds(S0, Clear(BlockA))
Holds(S0, Clear(BlockB))
Is it also true that *Holds(Result(Puton(BlockA, BlockB), S0), Red(BlockA))*?
That is, if Block A is red at the start, is it still red after we have put Block A on top of Block B? This question certainly is not difficult for a human to answer: of course Block A is still red. The problem is that this inference is not sanctioned by the theory. The theory as it stands says nothing about how a block's color is – or is not – affected by the occurrence of actions.

Thus, we have a problem: the problem of predicting – within the situation calculus – how things stay the same as actions occur. Now, this is not the frame problem, although researchers in AI have often identified this as the frame problem (see section 3). McCarthy and Hayes (1969) never named this problem; they just offered a way of handling it: namely, writing down axioms that specify how fluents do not change as actions occur. Such axioms are called *frame axioms*. An example of a frame axiom is:

$Holds(s,\ Red(b1)) \Rightarrow Holds(Result(Puton(b1,b2),\ s),\ Red(b1))$

or more generally:

$color(s,b1) = color(Result(Puton(b1,b2),s),\ b1)$

Such an axiom would allow the desired inference: that Block A is still red after putting it on Block B.

Of course, a tremendous number of things stay the same as Block A is moved to Block B: the president of the United States stays the same, congressmen are still in office, the Recommended Daily Allowance for zinc remains the same, and so on. To predict that all these fluents will remain unchanged during block moving, we would have to add a very large number of frame axioms to the theory. In general, if there are m actions and approximately n fluents that remain unchanged for each action, approximately mn axioms are needed – a vast number in any formalization of a reasonably complex world.

The need for such a large number of frame axioms in order to prove that most things stay the same as actions are performed is known as the frame problem. There are several reasons why the frame problem is, indeed, a problem. First, it is very time consuming and tedious to write down all the axioms that one would need to write in order to prove the desired facts. Second, such a system, bloated with obvious axioms, is inelegant. By the same token, it is counterintuitive; it seems unlikely that we reason that Block A is still red *because* we chain through a long list of axioms until we find one that says that the color of a block before a *Puton* action is the same as the color after the *Puton* action. Third, such a large number of axioms greatly slows down the system; the search for the proper axiom becomes time consuming. Fourth, and perhaps most seriously, it has been pointed out (McDermott, 1982) that frame axioms are often *false*, particularly in systems that allow for concurrent actions. If someone spray paints Block A as the *Puton* occurs, its color *will* change. Concurrent actions were (implicitly) not allowed within McCarthy's situation calculus, [3] but the broader problem remains.

3 What the Frame Problem Really Is and What the Frame Problem Was Taken to Be

In its most literal sense, the frame problem is the problem that arises from having a plethora of axioms in order to reason that most features about the world remain unchanged as actions occur. McCarthy and Hayes (1969), however, immediately identified the frame problem as the problem of predicting *within the situation calculus and without using frame axioms* that most things about the world remain the same as actions are performed. Very often, this problem is viewed as a representational one: How can we say in a concise manner, "Except for the features that are explicitly known to change, everything remains the same" ?

This is the classical frame problem. In fact, however, the frame problem has been reinterpreted in a variety of ways, all of them more general than this original formulation. We next present some of the major reinterpretations, going from the specific to the general.

On a slightly more general level, the frame problem has been identified with the *persistence problem* (Shoham, 1988): the general problem of predicting the properties that remain the same as actions are performed. Again, it is understood that this prediction takes place without frame axioms – but the prediction can take place within any reasonable formalism for reasoning about time – not just within the situation calculus. This may seem like a small change, but as we shall see, it is a critical one. Many solutions to the old situation calculus frame problem will not work for the more general persistence (or, as it is also known, the extended prediction) problem.

Going upward in the generality scale, the frame problem has been understood as encompassing both the persistence problem and its dual – namely, the problem of determining how things *change* over time. These two problems together are known as the (forward) *temporal projection problem* (Morgenstern and Stein, 1988).

[3] Some extensions to the situation calculus, such as that of Gelfond, Lifschitz, and Rabinov (1991), do allow concurrent actions, but they have not yet dealt with the frame problem.

It is interesting to note how the two parts of the temporal projection problem differ. The persistence problem is largely a representational problem: the idea is to predict that properties remain the same without a plethora of axioms explicitly stating that things do remain the same. In the case of predicting that things *change* over time, however, there is no objection to causal axioms stating that an action causes a particular effect to take place. On the other hand, there is a *computational* problem: figuring out everything that has changed when an action is performed can be a very time-consuming task. This is especially true if the world is very interconnected or if there are causal chains. For example, if I carry my briefcase into my office, then everything in the briefcase will also be in my office. So the result of my carrying action is not only that I and my briefcase are currently in my office, but also that my pens, my yellow pad, and today's *New York Times* are in my office. Likewise (Ginsberg and Smith, 1988), if I put a newspaper on top of an air vent, then my hand will be empty, and the newspaper will be lying on top of the vent. But in addition, the room may become stuffy (if that is the only air vent) or the pages of the newspaper may rustle (if the air flow is sufficiently powerful). The point here is that there are *derived* or *delayed* effects that are tedious to infer. It is not that the placing of the newspaper on the air vent directly causes the room to be stuffy. The placing of the newspaper causes the newspaper to block the vent; the vent being blocked then causes the room to be stuffy. [4] This problem is known as the *ramification problem* (Finger, 1988).

On a still more general level, the frame problem has been identified with the general problem of temporal reasoning. This includes forward temporal projection – reasoning about persistence, reasoning about change, and the ramification problem, as well as the *backward temporal projection problem* (Morgenstern and Stein, 1988): how can we reason about what happened at an earlier time point if we are told what is true at a later time

[4] Of course, one *could* explicitly write out all these derived and indirect causes. Such a strategy, however, is flawed for two reasons. First, it would be anti-intuitive: it is odd to say that the placing of the newspaper on the vent *causes* the room to be stuffy. Moreover, it would turn this from a computational problem to a representational problem, for we would then need a plethora of causal rules listing all the direct, indirect, and derived effects of an action. These casual rules would often be difficult to write down. For example: placing the newspaper on the air vent causes the room to be stuffy only if the air vent is the only source of air in the room.

point? Closely related to the problem of backward temporal projection is the problem of *explanation*: if we predict that a certain fact will be true at time t and we are wrong, can we explain what must have gone wrong previous to time t? It also includes the *qualification problem* (McCarthy, 1986): roughly speaking, the problem of specifying what conditions must be true in the world for a given action to have its intended effect. The classic example here is the case of turning the ignition key. If you turn the ignition key in your car, you expect the car to start. Actually, though, many conditions have to be true in order for this statement to be true: The battery must be alive, the starter must work, there must be gas in the tank, the tailpipe must be clear of obstacles, and so on. As with the original frame problem, this is a representational problem with a computational sidekick. We certainly would not wish to write down causal rules with cumbersomely long antecedents. Moreover, even if we were to write down such causal rules, would we necessarily know enough to apply them? I reason that I will be able to drive my car to work tomorrow morning, even though I do not know for certain that the battery will be alive and that there will be no potato in the tailpipe; in fact, I never explicitly consider these car parts. Finally, the problem includes general questions on the nature of causation: What is a causal rule? Do we even know if a causal rule is true? What is the connection between causation and material implication (the standard "if-then" connective of classical logic)? and so on (Shoham, 1988).

Lastly, some philosophers have interpreted the frame problem to be a general problem of reasoning. Fetzer (1991), for example, has argued that the frame problem is just an instance of the general principle of induction: we realize that Block A will be blue after the *Puton* action because that is the way it has been every other time we did the *Puton* action.

There are several points which must be made here. First, although the discussion above lists various interpretations of the frame problem from the most specific to the most general, the generalization takes place along more than one dimension. The move from the original frame problem to the persistence problem represents a generalization along the dimension of temporal ontology. The problem is the same, but it is examined within a broader theory of temporal reasoning. The move from the persistence problem to the forward temporal projection problem to the general problem of causal reasoning represents a generalization of the aspect of temporal reasoning that is being examined. However, these generalizations did not occur only in the manner and order presented in this discussion. For example, some researchers (e.g., Lifschitz, 1987) have closely combined a solution to the frame problem with a solution to the qualification problem, although they have addressed these problems only within the restrictive situation calculus.

Second, it can reasonably be argued that some of these generalizations – particularly the generalization to the problem of induction – are improper, false, or beside the point. If the frame problem can be generalized, as by Fetzer (1991), to the general problem of induction, then so can many other reasoning problems. One could in the same manner generalize probabilistic reasoning to the problem of induction: Of all the times that I have tossed a coin, it has landed on its head half the time, so I figure that if I toss a coin now, it has a 50 percent chance of landing on its head. But it is rather bizarre to think of probabilistic reasoning in this way. So too with the frame problem.

This brings us to the third point: What is the purpose of presenting these different interpretations of the frame problem in increasing generality? It is the contention of this chapter that many of the problems with solutions to the frame problem have occurred precisely because they addressed the problem at the wrong level of generality. Solutions that are too specific often miss the salient part of the frame problem – as discussed in section 5. On the other hand, attempts at a solution to too broad a problem have typically remained only attempts, not actual solutions, precisely because they are so difficult.

Just where should research be aimed? My present feeling is that it is wisest to identify the frame problem as the persistence problem, but at the very least, one should work on the general problem of forward temporal projection. We can state a very general dictum: all other things being equal, the more general the solution, the better. But usually all other things are not equal. Typically, solutions do not differ solely in their level of generality. Moreover, if solutions A and B both solve the frame problem, and only solution B solves the qualification problem, but A is a better solution than B, I would still argue that A should be preferred. In any case, whatever the level of generality, a solution that is offered at one level should not be *inconsistent* with our intuitions of a solution at a higher level.

The proper degree of generality is just one of the properties that a good solution to the frame problem must have. The next section presents a list of such properties.

4 Requirements for a Solution to the Frame Problem

It is easy to cavil about existing solutions to the frame problem and to find for each some point(s) of objection. To be fair, however, we should approach each solution with a good ides of what our minimum requirements of a solution are, and then analyze each solution with respect to these requirements.

We list such requirements below. Because this is intended to serve as the minimum list, it is probably not comprehensive.

[1] **Achieving the right level of generality/specificity.**
As discussed in the previous section, the frame problem has been viewed as a very broad problem of induction, a very specific problem in temporal reasoning, or somewhere in between. Although we do not argue for a particular level of specificity, we do place the following two constraints on generality/specificity:

a. The solution must go beyond overly restrictive temporal languages such as the situation calculus. In particular, a solution should work for an expressive temporal reasoning system that allows for concurrent actions, the representation of "gaps" (i.e., periods where one does not have complete knowledge of all that is going on in the world) and partial specification of actions (i.e., describing an action such as going down to Hertz, renting a Ford, getting a map from the rental agency, and driving to Buffalo).

b. The solution should be compatible with a general theory of temporal reasoning. This does not mean that the solution should necessarily *solve* the qualification or ramification problem, or that it should give a convincing account of causation. But it does mean that the solution should not *preclude* further work in causal reasoning. For example, as discussed in Section 5.3.3, the works of Kautz(1986), Lifschitz(1986), and Shoham (1988) are inconsistent with a theory that allows for backward temporal reasoning. These solutions are thus not satisfactory for our purposes.

These constraints allow a wide range of solutions of varying degrees of generality. Within this acceptable range, however, the following dictum applies: all other things being equal, the more general the solution, the better.

[2] **Staying faithful to intuitions.**

Following one's intuitions is in general a wise principle to follow when developing AI theories. It is certainly true in the case of developing a solution to the frame problem.

This probably means incorporating some notion of causation into the theory. In particular, the theory should be based on the following principles:

a. The performance of an action causes certain things about the world to change.

b. Most other features about the world stay the same.

c. Principle b. is not an absolute rule. It is typically true but can sometimes be false. In general, the more concurrency is allowed (i.e., the more things happen in the world at the same time), the greater the likelihood that Principle b. will be false.

Ignoring one's intuitions often means that problems with one's solution crop up later. Examples of this are described in section 5.

[3] **Truth and soundness.**

The axioms in one's theory should be true. Approximations and/or misstatements may be convenient, but a false theory will eventually lead to incorrect results. (Of course, all theories are at best approximations of the real world, and thus may be considered false to some extent. It is a question of threshold, and it is up to AI researchers to get this threshold right.)

[4] **Concreteness.**

The theory should demonstrate actual solutions to some set of concrete benchmark problems. Solving concrete problems is essential; it forces us to be intellectually honest and to work out things in detail. It points out all sorts of difficulties that did not exist at a higher level of abstraction. Solving a concrete non-toy problem is usually tedious and sometimes exceedingly hard. Thus, researchers who have examined concrete non-toy problems often do so in very narrow domains. This may explain why there is an inverse relationship between the generality of a problem and the concreteness of the solution. As a rule, researchers who study the more general and all-encompassing problems in temporal reasoning do not offer concrete solutions.

[5] **Concise representation.**

This means, specifically, no endless list of frame axioms or rules of persistence (a few general principles of persistence are okay). This may seem like an obvious criterion – we are, after all, discussing a solution to the frame problem, which is the problem of having too many frame axioms. In fact, it is not all that obvious, and it is not always satisfied. Some of the theories that this chapter examines were developed as theories of (some aspects of) temporal reasoning, and do deal with problems of persistence and causation. They may be far superior to McCarthy and Hayes's situation calculus in many respects, but still rely on a large number of persistence rules to get the proper inferences. These may be quite good theories of causal reasoning in some ways, but they are not solutions to the frame problem.

[6] **Based on theory.**

Some of the solutions that we examine here are procedural. Procedural solutions are fine, as long as they are strictly based on some theory. This allows for formal and rigorous reasoning about and comparison among these systems, and for analyses relative to the points above.

[7] **Simplicity.**

Like generality, this is a relative preference criterion. In general, all other things being equal, the simpler and more elegant the solution, the better. Occam's razor, in other words.

5 How Solutions Measure Up

The previous section listed some of the minimum requirements that any solution to the frame problem should satisfy. Although these criteria seem rather simple, in practice virtually none of the available solutions measure up against this list. This section presents some major trends in research on the frame problem and analyzes the solutions in light of the previous section.

This is not a comprehensive survey of all previous research on the frame problem. Rather, we aim to identify the major types of solutions. We then wish to see in which areas they fall short, what the underlying philosophical justifications are for these solutions, and how it is possible to set research on the right course. Many classifications of solutions are possible. We choose to identify and discuss the following trends: the monotonic, the procedural, the nonmonotonic, and the probabilistic/statistical. Since the lion's share of recent research on the frame problem has been in nonmonotonic reasoning, the section on nonmonotonic solutions is the longest. It is divided into several sections: naive nonmonotonic temporal reasoning, forward reasoning approaches, causal approaches, progressive approaches, and current trends.

5.1 Monotonic Approaches to the Frame Problem

Any solution to the temporal reasoning problem within monotonic logic must include some frame axioms that specify how the world stays the same as actions are performed. The aim of any monotonic solution to the frame problem is to somehow state the theory in such a way that the number of frame axioms is kept at a reasonable level. Two classes of solutions have been proposed.

One solution, *framing primitive fluents by events* (Davis, 1990), asserts within the logic that only certain specified states and fluents are changed by an event, and that all others remain the same. To accomplish this, one designates a few states and fluents as *primitive*; the rest are *derived*. The frame axioms assert that only specified primitive fluents and states change during the event; the rest remain the same.

The advantage of this representation is that we need only one axiom for each action type (plus one axiom for each primitive state), resulting in a very manageable number of frame axioms. However, Davis himself pointed out some of the disadvantages: First, concurrent actions are not allowed in this representation. Second, the distinction between primitive and derived states is artificial and conflicts with the spirit of the logicist approach.

The second approach, *explanation closure*, was suggested independently by Davis (1990) and Schubert (1990). (Schubert generalized on a simpler version of an idea presented by Haas,1987.) The idea is that the only way a fluent can change is if a certain (primitive) event happens. Equivalently, if the event does not happen, the fluent remains unchanged. For example, one could say that the color of a block remains unchanged unless a paint action occurs. If one maintains a distinction between primitive and derived events, as Davis does, this approach uses about the same number of axioms as the first approach. Otherwise, one needs about $2 \times \mathcal{F}$ axioms, where $\mathcal{F}$ is the number of fluents. (Schubert calls these axioms *explanation closure axioms*, because they give a complete explanation of how a fluent changes.) Reiter (1991) suggests a modification of this approach which brings down the number of axioms to $\mathcal{F} + \mathcal{A}$, where $\mathcal{A}$ is the number of actions.

This approach uses a reasonable number of axioms and has the advantage of allowing concurrent events. Nonetheless, there are two major disadvantages: First, while concurrent actions are allowed if they are known, these systems cannot handle *unknown* concurrent actions. One cannot reason correctly in a world in which not everything is known. That is, the system's knowledge about the events in the world must be complete. But this is hardly a realistic assumption. Moreover, it is not needed in typical common-sense reasoning. I reason that Block A will be blue after I move it even though I do not know everything else that is happening. I reason that Rudy Giuliani is still mayor of New York when I wake up, even though I do not know most of what has happened during the night.

Second, one must add very strong assumptions in order to make very straightforward inferences in this system. For example, consider a simple blocks world theory with axioms saying that blocks A, B, and C are on the table and are colored blue in situation S0, and that Block B is placed on Block C, resulting in situation S1. To infer that Block B is still blue in S1, or that Block A is still on the table in S1, one would have to add axioms that either specify that no other actions happen or that severely restrict the types of action that happen. It may be fine to conclude these facts using plausible inference (see section 5.3.5 for a description of MAT, which allows just this sort of plausible inference), but adding such axioms to a monotonic logic is problematic. Such assumptions are typically not true in the world (where many things may happen as Block B is being moved to Block C) and are not needed by people who do common-sense reasoning.

Similar problems beset other attempts to solve the frame problem within a monotonic logic. For example, Elkan's (1992) logic cannot handle concurrent actions. Moreover, it requires very strong assumptions in order to work.

5.2 Procedural Approaches

Procedural solutions to the frame problem have been around since the earliest days of the frame problem. The best known of such approaches is STRIPS (Fikes and Nilsson, 1971). STRIPS is first and foremost a planning program: Given an initial state, a goal state, and a list of actions, it will plan a sequence of actions that achieves the goal state, using a planning method known as means-end reduction. In the course of planning, however, it must reason about what changes and what stays the same after an action is performed. If one starts out with blocks A, B, and C on the table, and one plans to build a tower with A on top of B and B on top of C, a possible plan is performing *Puton(B,C)*, followed by *Puton(A,B)*. But to reason that this plan is feasible, one must know that block A will still be on the table after B is placed on C. If A has suddenly disappeared, the plan is no longer feasible. Thus, STRIPS must deal with the frame problem.

STRIPS solves the frame problem by assuming that if an action is not specifically known to change some feature of the world, it does not. This principle is easy to represent procedurally (thus the popularity of procedural solutions to the frame problem). STRIPS works as follows:

Each state is represented as a set of statements, describing the world at a particular time. For example, the initial state might be:

{ *clear(A), on(A, Table), clear(B), on(B, Table), clear(C), on(C, Table)* }

and the goal state might be:

{ *on(A,B), on(B,C)* }

Associated with each action is a list of preconditions that must be satisfied in order for the action to be performed, along with an *add list* and a *delete list*. The add list is the set of statements that gets added to the current state after an action is performed, and the delete list is the set of statements that gets deleted from the current state after the action is performed. For example, *Puton(x,y)* could have the associated precondition list: { *on(x,z), clear(x), clear(y)* } , add list: { *on(x,y)* } , and delete list: { *clear(y), on(x,z)* } .

Let *s0* be the current situation and let *S0* be the set of statements representing that situation. Then, if $s1 = result(act,s0)$, $S1 = S0 \cup addlist(act) - deletelist(act))$.

The advantages of STRIPS are clear: it is computationally efficient and easy to understand. Proponents argue that STRIPS reflects the common-sense reasoning that people do: after all, we usually just figure out what has changed and ignore the rest. However, there are severe disadvantages to STRIPS. By its very nature, it cannot handle conditional actions (Pednault, 1991; Penberthy, 1993). More importantly, STRIPS works only for limited temporal ontologies. In particular, it will not work if unknown concurrent actions are allowed. This is because the assumption underlying STRIPS – "if it isn't explicitly declared to change, it doesn't change" – is simply false in a world in which unknown concurrent actions are allowed. If someone grabs A at the same time as I move B to C, Block A will *not* be on the table at the end of my action! STRIPS is incapable of representing or reasoning about such a scenario. STRIPS thus violates criterion [1] of our minimum requirements cited earlier.

STRIPS fails for the same reason that frame axioms fail: because they are false, particularly in any world that allows for concurrent actions. The fact is that while most features about the world usually do not change during an action, some do. It is true that moving a block does not change the president of the United States, but if I move a block while a new president takes the oath of office, the president after my action will not be the same as the president before my action. The world's population is changing every minute, so the world's population after an action will always be different than the world's population before the action, even though the action is not *causing* the population of the world to change.

The point, then, is to somehow capture the fact that *most* features usually do not change as actions are performed. Vague words like *usually* and *most* are difficult to capture within classical logics. Thus many researchers have looked elsewhere – in particular to nonmonotonic and probabilistic logics.

5.3 Nonmonotonic Approaches

5.3.1 Origin of Nonmonotonic Approaches

In section 3, we argued that the classical frame problem can be viewed as the problem of saying in a concise, formal manner, "Except for the features that are explicitly known to change, everything remains the same." That is, we would like to achieve the following in some formal system: Assume that a feature remains the same after an action is performed, unless it is explicitly known to change.

This sort of statement cannot be formalized in a standard, classical logic. In such logics, one cannot explicitly talk about assumptions, exceptions, or what usually happens. Since such concepts are prevalent in common-sense reasoning, AI researchers began in the 1960s and 1970s to develop logics that could handle such concepts. The result of this research was the development of a family of logics known as *nonmonotonic logics*. These logics are designed to formally capture concepts such as *usually*, *typically*, and *most*. The canonical example, solved by all nonmonotonic logics, is that of Tweety the bird. Given (only) the assumptions that birds typically fly and that Tweety is a bird, a nonmonotonic logic will support the conclusion that Tweety flies.

These logics are called nonmonotonic in contrast to classical logics, where the set of conclusions is monotonic with respect to the set of assumptions: the more assumptions you have, the more conclusions you can draw. In nonmonotonic logics, you may have to *retract* a conclusion as new assumptions are added. For example, if you add to the above set of assumptions the fact that Tweety is an ostrich, and you know that ostriches do not fly, you must retract the conclusion that Tweety can fly.

There are various forms of nonmonotonic logic; one of the best known of these is *Circumscription* (McCarthy, 1980). The idea behind Circumscription is to restrict the extension of a predicate (or set of predicates) as much as possible; that is, to limit the entities that satisfy a certain property. For example, we might represent the statement "birds typically fly" as "birds fly unless they are abnormal in some respect" :

$\forall\ x\ bird(x) \wedge \neg\ ab(x) \Rightarrow fly(x)$

We would then wish to restrict the class of abnormal birds as much as possible. Penguins and ostriches and birds with broken wings would be abnormal birds, but if a bird could not be shown to be abnormal, one would conclude that it was *not* abnormal, and thus, could fly.

McCarthy proposed that this idea could be extended to temporal reasoning. Specifically, he suggested that one could formulate the principle of inertia – i.e., that most things do not change – by adding the following axiom to one's theory:

$\forall\ f,e,s \quad Holds(f,s) \wedge \neg\ ab(f,e,s) \Rightarrow Holds(f,Result(e,s)).$ [5]

That is, if a fluent holds in a particular situation, and an event occurs that is not abnormal with respect to this fluent in this situation, then the fluent will still be true in the situation

[5] I am stating here Hanks' and McDermott's (1986) restatement of McCarthy's approach. McCarthy had one statement of inertia for each fluent. The generalization to one principle of inertia is straightforward. Note also the reversal of the fluent and situation arguments in the *Holds* predicate; both notations are common in AI.

resulting from performing that event.

For example, consider the action *Move(A,B)*. The *Move* action is not abnormal with respect to the president of the United States, the color of my carpet, and the color of blocks A and B. It is abnormal, however, with respect to the location of blocks A and B. Thus, if Block A is blue before A is moved onto B, it will be blue afterwards. Likewise, if Bush is president before the block is moved, he will be president afterwards. On the other hand, the location of Block A will have changed.

Has McCarthy, in fact, solved the frame problem? As discussed in the next section, McCarthy's solution has a major flaw – it succumbs to the Yale Shooting Problem, a special case of the multiple extension problem. Moreover, I contend that even if it were not for this flaw, McCarthy's solution would still not be an adequate solution to the frame problem. McCarthy's solution is so strongly based on the situation calculus that it cannot be extended to a more expressive temporal ontology. In particular, it cannot be extended to any temporal ontology rich enough to handle concurrency. For if we closely examine the principle of inertia described earlier, we see that it is *false* if (unknown) concurrent actions are allowed. The principle of inertia says that the *Move* action is not abnormal with respect to the color of the blocks – but what if someone spray paints Block A at the same time as someone else moves it? Or suppose that the *Move* action takes place at the moment that Bill Clinton is sworn into office. The *Move* action is not abnormal with respect to the presidency – yet Bush is president in the situation before the *Move* action, and Clinton is president in the situation resulting from the *Move* action. The principle that McCarthy has formalized is false, precisely because concurrent actions are not taken into account. [6]

Thus, McCarthy's solution fails in two respects:

(1) It succumbs to the Yale Shooting Problem (to be discussed at length in the next section).

(2) It does not work for concurrent actions.

As we shall see in the next section, much has been made of problem (1), and much energy has been expended in fixing it. But virtually all of these solutions ignore problem (2), and thus are still inadequate as solutions to the frame problem. Ironically, what problem (1) – the Yale Shooting Problem – has achieved has been to divert the focus of research on the

[6] Note that this flaw also troubles the STRIPS procedural solution to the frame problem (section 5.2).

frame problem from the honest analysis of the advantages and disadvantages of a particular approach to the plugging of a hole in a solution that is not very good to start with.

5.3.2 The Yale Shooting Problem

McCarthy's solution was for a short time believed to have successfully solved the frame problem. In 1986, however, Hanks and McDermott showed that McCarthy's solution succumbs to a serious problem, popularly known as the Yale Shooting Problem. The Yale Shooting Problem can itself be seen as an instance of the multiple extension problem – whose existence has been know since the inception of nonmonotonic logics.

The *multiple extension problem* is, briefly, the problem that arises when two default rules conflict. The canonical example is the Nixon-Quaker-Republican diamond (Reiter and Criscuolo, 1981): It is well-known that Quakers are usually pacifists and that Republicans are usually nonpacifists. Suppose we also know that Nixon was a Quaker and that he was a Republican. Was Nixon a pacifist or a nonpacifist? There is support for both positions, so it is difficult to conclude anything about Nixon's beliefs. This is the multiple extension problem (more than one model, or extension, is supported by the facts); but indeed, it originally was seen as a *feature* of nonmonotonic logics. The multiple extension problem showed that one could represent conflicting information within a nonmonotonic theory. In fact, such conflicts often occur in common-sense reasoning. We often have conflicting default information with regard to a particular subject. Sometimes we can weigh the conflicting information and come to a decision, and sometimes we cannot. This is all consistent with common-sense reasoning.

In the Nixon diamond, it is not clear that we would *want* to make a decision vis-à-vis Nixon's pacifism given only the above information. So the fact that we have multiple extensions is not in itself an anomaly. The problem arises in situations where we do have a definite intuition about what conclusion a set of axioms should support. This is the case of the Yale Shooting Problem.

The Yale Shooting Problem was discovered when Hanks and McDermott (1986) attempted to integrate temporal and nonmonotonic logics. It can be briefly described as follows: We are told that a gun is loaded at time 1, and that the gun is fired at Fred at

time 5. Loading a gun causes the gun to be loaded, and firing a loaded gun at an individual causes the person to be dead. In addition, the fluents *alive* and *loaded* persist as long as possible; that is, these fluents stay true unless an action that is abnormal with respect to these fluents occurs. Thus, a person who is alive tends to remain alive, and a gun that is loaded tends to remain loaded. What can we conclude about Fred's status at time 6? If we work within the situation calculus and we assume that the starting situation is *S0*, we can phrase the following question: Is
Holds(Alive(Fred), Result(Shoot, Result(Wait, Result(Wait, Result (Wait, Result(Load, S0)))))) true? [7] Although common sense argues that Fred is dead at time 6, the facts support two models. In one model (the expected model), the fluent *loaded* persists as long as possible. Therefore, the gun remains loaded until it is fired at Fred, and Fred dies. In this model, at time 5, shooting is abnormal with respect to Fred's being alive. In the other, unexpected model, the fluent *alive* persists as long as possible (i.e., Fred is alive after the shooting). Therefore, the fluent *loaded* did not persist; somehow the gun must have become unloaded. That is, in some situation between 2 and 5, *Wait* was abnormal with respect to the gun being loaded.

Here, the existence of multiple extensions is a genuine problem. If, given the simple set of assumptions above, one cannot even conclude that the gun stays loaded and that Fred is dead, in what sense can we say that McCarthy has solved the frame problem? Hanks and McDermott in fact argued that the existence of the Yale Shooting Problem underscored the inadequacy of logicist temporal reasoning. Other researchers viewed the YSP as just one more challenge to be solved.

Although many solutions have been proposed to the Yale Shooting Problem, they can be grouped into a few broad categories. The two best known of these are the chronological and causal-based approaches. We discuss these, as well as an alternate approach, below. Our interest in examining these approaches is exclusively to determine whether they are adequate solutions to the frame problem.

[7] The Yale Shooting Problem was originally formulated within the situation calculus. Most of the proposed solutions to the Yale Shooting Problem have been given within the situation calculus, a fact that we believe helps explain the lack of progress in research on the frame problem. But the YSP is not unique to the situation calculus and can exist in any temporal formalism (McDermott, personal communication, 1988).

5.3.3 Chronological Approaches to the Yale Shooting Problem

The first solutions to the Yale Shooting Problem (Kautz, 1986; Lifschitz, 1986; Shoham, 1988) work by imposing a forward-in-time order on reasoning. They are thus known as the *chronological approaches.* In the expected model, one reasons from the earliest to the latest time point; in the unexpected model, one reasons from the latest time point (after the shooting) to earlier times. Thus, the unexpected models are disqualified, and we can conclude that Fred is dead after the shooting.

Although these solutions do work for the Yale Shooting Problem, it is clear that they do not solve the frame problem. First, the strong constraint on forward-in-time reasoning works well when one is considering problems of *prediction* (i.e., will Fred be dead after the gun is fired?), but does not work when one is considering problems of belief revision or explanation. In such cases, backward-in-time reasoning is necessary. For example (Kautz, 1986), suppose there is a default rule that states that a car will typically stay where it is parked. If I park my car in a lot at 9:00 a.m., I will predict that the car will be there when I return at 5:00 p.m. But if I return to find the car gone, forward-in-time reasoning will entail that the car disappeared from the lot at the last minute – just before 5:00 p.m. ! This is clearly not a reasonable conclusion. Chronological approaches are thus inadequate for the general problem of temporal projection. Second, Lifschitz's and Kautz's solutions are firmly based on the situation calculus, and thus inherit all the problems that the situation calculus entailed for McCarthy's original solution.

5.3.4 Causal-based Approaches

Chronological approaches, popular for a short time after the publication of the Yale Shooting Problem, quickly lost favor – due for the most part to the fact that they are incapable of handling backward temporal reasoning. The next wave of proposed solutions, which we refer to as *causal-based solutions* (Lifschitz, 1987; Haugh, 1987; Baker, 1991), differed from the chronological approaches in two major respects: (a) there is an explicit attempt to make sure that these solutions work for both forward and backward reasoning and (b) there is an explicit attempt to base the solutions to the Yale Shooting Problem on a strong intuition about temporal reasoning.

The principle behind the chronological approach – that changes should happen as late as possible – seemed to work, at least at first glance. But this principle does not seem intuitive, or even true, and it does not seem to have anything to do with our understanding of why the gun should remain loaded and Fred should die. The causal-based approach argues that we expect Fred to die because there is an action that causes Fred's death, but there is no action that causes the gun to become unloaded. All the causal-based approaches formalize this principle in some way. Below we examine Lifschitz's (1987) approach (which also solves the qualification problem).

Lifschitz defines a predicate *affect(a,f,s)* where an action a affects a fluent f in some situation s if the action is successful and causes the fluent to take on a particular truth value (all variables are assumed to be universally quantified unless otherwise noted):

$affects(a,f,s) \Leftrightarrow success(a,s) \wedge \exists v \;\; causes(a,f,v)$

An action a is successful in a situation s if all of its preconditions hold in that situation:

$success(a,s) \Leftrightarrow \forall f \;\; (precond(f,a) \Rightarrow holds(f,s))$

There is also an axiom describing how fluents change:

$success(a,s) \wedge causes(a,f,v) \Rightarrow (holds(f,result(a,s)) \Leftrightarrow v = true)$

The idea then is to minimize the predicates *precond* and *causes*. Minimizing *precond* solves the qualification problem (i.e., the only qualifications to an action are those which are explicitly stated or which can be derived); minimizing *causes*, claimed Lifschitz, solves the frame problem. This theory correctly handles the Yale Shooting Problem. Since there is no axiom that says that the *Wait* action causes the gun to become unloaded, the minimization of *causes* will entail that the gun cannot become unloaded in the Yale Shooting Scenario.

A solution similar to Lifschitz's was independently suggested by Haugh (1987). Moreover, researchers have suggested several modifications of Lifschitz's solution (Lifschitz and Rabinov, 1989; Baker, 1991) to correct some of the anomalies in Lifschitz's original solution. Much has been made of the anomalies, which we briefly discuss here, but as in the case of the Yale Shooting Problem in general, we believe that the emphasis on the anomalies has blinded most of the AI community to the real deficits of this entire approach. Shortly after Lifschitz proposed his approach, it was discovered that although his theory could handle backward reasoning, it did so a little strangely. For example, if one finds out at time 6 that

Fred is alive, the theory would entail that the action *Wait* always causes the gun to become unloaded. This would mean that the Yale Shooting Scenario would *never* work; the moment someone would wait between a loading and a shooting, the gun would become unloaded. Lifschitz and Rabinov modified the theory to include the existence of one-time "miracles" so that in cases of unexpected occurrences one can assume that a "miracle" happened. Thus, if one finds out at time 6 that Fred is alive, one can assume that a miracle must have unloaded the gun, rather than permanently alter one's causal theory. Baker (1989) noted that Lifschitz's theory did not work correctly for some temporal reasoning involving a combination of forward and backward reasoning (most notably the Stanford Murder Mystery, also discussed in Baker, 1991), and suggested ways of modifying Lifschitz's theory so that these problems could be handled.

The discovery of various holes in Lifschitz's theory, the attempts to plug these holes, and the subsequent discovery of still more problems have led many (e.g, (Elkan, 1992)) to argue that these solutions to the Yale Shooting Problem cannot serve as solutions to the more general problem of temporal reasoning; in particular, that these solutions work only for the few toy problems for which they are developed.

We agree with this conclusion: that the causal-based approach has not yielded a satisfactory solution – but not only because each new iteration of the approach seems to fall prey to some new twiddle on the Yale Shooting Problem. More importantly, the causal-based approach fails in two important respects: First, it does not work for concurrent or unknown actions. Second, it is not consistent with our intuitions.

The causal-based approach does not work for systems allowing concurrent or unknown actions because it is based on the situation calculus. In particular, it relies on the very strong assumptions of the situation calculus – that only one action happens at a time and that all actions are known. (This is beginning to sound like a refrain of this chapter – and indeed it is. Given that concurrent and unknown actions are so commonplace in commonsense reasoning, it is amazing how few theories of temporal reasoning have ever considered systems that allow for such actions.) If we add the axiom *causes(unload,load,false)* to the Yale Shooting Problem, and allow for concurrent actions, the conclusion that Fred is dead at time 6 no longer follows. For indeed, there is a model in which an *Unload* action happens at the same time as the *Wait* action, the gun is unloaded when the *Shoot* takes place, and Fred remains alive. The sole reason that the causal-based approach works for the Yale Shooting Problem is that no other actions are allowed to happen during the *Wait* action.

Now, this realization may be somewhat hard to take. Wasn't the point of the causal-based approach to formalize a very intuitive principle of temporal reasoning – that there are no unexpected causes? Shouldn't this principle be enough to ensure that a surprise *Unload* action doesn't happen – without resorting to a complete ban on concurrent actions? And if not, what has the causal-based approach bought for us anyway?

The answers to these questions are, respectively: so we thought, not really, and it bought us something, but not all that we expected. The frame problem was originally discovered in the very restrictive situation calculus. Even in a system in which concurrent actions were not allowed, it was very difficult to infer that a red block was still red after a move occurred. To make such an inference within a standard, monotonic logic, frame axioms were needed. Frame axioms were the only way to ensure that actions did not have strange effects (i.e., that a *Move* action did not cause a red block to turn blue) or that fluents did not change values by themselves. What Lifschitz's theory (and those of his successors) achieved was to allow the correct inferences without relying on frame axioms. These theories do concisely formalize the principle: A fluent will change only if there is some known cause.

However, that is about all these theories do. They do not, in and of themselves, preclude unexpected actions (such as the unloading of a gun) from happening. It is the restrictive situation calculus that does that. Take any of these theories away from the framework of the situation calculus, and they crumble entirely. Thus, these theories do not provide satisfactory solutions to the frame problem.

5.3.5 A Progressive Approach

Interestingly, the causal-based approaches described above do not even provide intuitively satisfying solutions to the *original* Yale Shooting Problem. For our intuition that Fred is dead at time 6 is not predicated on the strict situation calculus assumption that concurrent actions are not allowed. The natural reading of the story – a gun is loaded at time 1, and the gun is fired at Fred at time 5 – allows *anything* to happen during the passage of time.

The point is that there is no reason for an *Unload* to occur. Thus, we do not expect it, and models in which an *Unload* does occur are less preferable than models in which the gun stays loaded.

The moral behind the Yale Shooting Problem seems to be that we prefer models in which unexpected actions do not happen. The formalization of this principle has led to a third approach to solving the Yale Shooting Problem, outlined below and described in (Morgenstern and Stein, 1988; Stein and Morgenstern, 1994). Unlike most previous solutions to the Yale Shooting Problem, Motivated Action Theory (MAT) was not based on the situation calculus. Instead, it was based on a simple, interval-based theory of time, taken from McDermott (1982). Concurrent actions were allowed, as were unknown actions and gaps.

A description of a problem scenario in MAT is known as a *theory instantiation*, consisting of a theory T and a (partial) chronicle description CD . Intuitively, T gives the rules governing the world's behavior and contains causal rules and persistence rules. *Causal rules* describe how actions change the world; *persistence rules* describe how fluents remain the same over time. CD describes some of the facts that are true and the actions that occur during a particular interval of time.

Central to MAT is the concept of motivation. Intuitively, an action is *motivated* with respect to a theory instantiation if there is a "reason" for it to happen. The most important types of motivation are *strong* and *weak* motivation. An action is *strongly motivated* if it "has to happen" in all models (i.e., it is a theorem that the action happens). It is *weakly motivated* if it "has to happen" with respect to a particular model (i.e., if it must occur given the particular way the model is set up). Formally, a statement γ (saying that some action has occurred) is weakly motivated in a model if there is a sentence of the form $\alpha \wedge \beta \Rightarrow \gamma$ in the theory; α is motivated in the model, and β is true in the model. Thus, the effects of causal chains are weakly motivated. A model is preferred if it has as few unmotivated actions as possible. This allows for a very nice solution to the Yale Shooting Problem. In the expected model, where the gun remains loaded and Fred dies, there are no unmotivated actions. In the unexpected model, there is an unmotivated action – the *Unload* action.

Thus we prefer the expected models over the unexpected models. [8]

MAT has many advantages over earlier theories of temporal reasoning. It allows for forward and backward temporal reasoning, supports a flexible temporal ontology, is based on sound intuitions, and has been demonstrated to work on a large set of concrete benchmark problems. Despite all these advantages, however, it has been argued that it too falls short as a solution to the frame problem. For central to MAT's success are its persistence rules – which are just another form of frame axioms.

There are two responses to this objection. First, we do not have a plethora of persistence rules; we merely need one persistence rule for each fluent type. In fact, MAT uses no more persistence rules than Davis's solution of *framing primitive events by fluents*, with none of that solution's disadvantages. However, it might be argued that even this many persistence rules are too much, and that MAT has not provided a total solution to the frame problem. Note that while Lifschitz and his successors' theories use nonmonotonic logic to obviate the need for frame axioms and to show that the actions that occur have no unexpected effects, MAT uses nonmonotonic logic so that we can infer that no unexpected actions have happened. Both theories attack different aspects of the temporal reasoning problem, but no solution by itself is a fully adequate solution to the general frame problem. No theory satisfies all the criteria outlined in Section 4.

Since we have two solutions which solve different aspects of the frame problem, a natural question arises: Can these solutions be integrated? If so, does the resulting theory solve the frame problem?

It is difficult to add MAT to Lifschitz's theory since Lifschitz's theory is so strongly based on the situation calculus. However, one can add the basic idea behind Lifschitz's theory – minimizing the predicate *causes* – to MAT. This approach is briefly discussed in (Stein and Morgenstern, 1994), where we suggest performing two stages of minimization: first circumscribing the *causes* predicate, and then using MAT's model preference criterion on the resulting set of models.

Indeed, there has recently been a trend toward integrating existing solutions into new

[8] Note that minimizing unmotivated actions is not the same as minimizing action occurrences, particularly in cases where we have causal chains of action.

general purpose theories of temporal reasoning. We turn our attention to this approach next.

5.3.6 Current Trends: Integration into General Temporal Theories

The first wave of euphoria, or at least relief, following the first set of solutions proposed to the Yale Shooting Problem quickly ebbed as these solutions were shown to be inadequate for slight modifications to the YSP. As the cycle continued – new solutions were suggested and counterexamples discovered for these solutions – dissatisfaction with this ad hoc method of research grew.

Researchers have responded in two ways. On the one hand, they have attempted to lay out a set of benchmark problems that would touch on all the "hard" problems of temporal reasoning (Lifschitz, 1988; Sandewall, 1994, chap. 7). At the same time, however, they have tried to distance themselves from individual benchmark problems. Instead of constructing a theory and demonstrating that this theory could solve a set of particular problems, they have proven general properties of such theories. Thus, researchers have attempted to show that a temporal theory solves a certain *class* of temporal reasoning problems. The fact that this theory solves a *particular* benchmark problem is then a consequence of the more general theorem that has been proven.

Early examples of such work can be found in (Lin and Shoham, 1991; Lifschitz, 1991). Lin and Shoham focused on constructing theories which have a property that they label *epistemological completeness.* This property entails that if one is given a complete description of the initial situation, including the actions performed in that situation, one can completely predict the situation resulting from the performance of the actions. Thus, if a theory can be shown to be epistemologically complete, it can at least handle forward temporal projection.

In a similar but broader spirit, Lifschitz constructed a general theory of action which embedded Baker's (1989) approach to temporal reasoning. Lifschitz's main innovation lay in the fact that he proved that his theory would correctly handle temporal projection and ramification for a restricted class of temporal reasoning problems. Moreover, by embedding an existing theory into a more general theory of temporal reasoning, Lifschitz sparked another important and welcome trend: integrating and building upon existing theories, rather than starting each new temporal theory from scratch. Both these trends are evident in the work of Kartha (1993) as well.

The most ambitious of these projects to date has been that of Sandewall (1994), who has set out to assess virtually all major nonmonotonic temporal formalisms within a single unifying framework. Sandewall has painstakingly defined a taxonomy of temporal reasoning languages and has made explicit the characteristics – such as non-determinism, ramifications, and concurrency – that apply to, or are part of, each language. [9] He has also made explicit the many epistemological assumptions underlying benchmark problems and theories of temporal reasoning. Finally, he has mapped various theories of nonmonotonic temporal reasoning into his framework, and he has proven general results about these theories. For example, Sandewall has shown that Kautz's (1986) approach to nonmonotonic temporal reasoning (a type of chronological minimization) works correctly for the class of languages with the characteristics of inertia and strongly deterministic equidurational change, as long as certain epistemological properties hold: namely, the initial situation is completely known, and nothing is known about any subsequent situation.

To what extent have the current trends contributed toward a solution to the frame problem? Certainly they have injected a fresh spirit of enthusiasm into the research community. In many respects, however, they have fallen short of their promise. It is no doubt preferable to prove that a theory of temporal reasoning works correctly for a class of problems than to show that it works for specific benchmark problems. But ultimately such results are of interest only if they apply to broad and useful classes of problems. Unfortunately, in this respect, the results that have thus far been demonstrated are disappointing, for they apply only to very narrow sets of problems. For example, Lin and Shoham's (1991) epistemologically correct theories deal only with scenarios in which everything is known about the initial situation; thus they cannot handle problems like the Stanford Murder Mystery. Their theory was extended to deal with concurrent actions (Lin and Shoham, 1992), but only known concurrent actions are considered; moreover, the epistemological assumptions remain very

[9] A *characteristic* typically applies to or is part of a language if that language is expressive enough to describe the characteristic or contains instances of that characteristic. For example, ramifications are part of a temporal language if there are some causal laws that allow for indirect effects.

strong. Lifschitz and Kartha likewise assume that all actions are known (although they at least allow for the possibility that not all fluents are known in the initial situation); in addition, concurrent actions are not considered.

In contrast, Sandewall explicitly discusses broader classes of theories in which non-determinacy, ramifications, and concurrency are allowed, and which have very weak epistemological assumptions. At this point, however, Sandewall's *results* are limited to narrow classes of theories; it is assumed that all actions are known, and concurrent acts and ramifications are not considered. Thus, none of his results as yet represent an adequate solution to the frame problem. Sandewall plans to address broader classes of theories in the sequel to his 1994 work. Correctness results for a sufficiently broad class of theories would represent genuine progress toward solving the frame problem within a nonmonotonic temporal formalism. At the very least, Sandewall has made explicit the many assumptions underlying existing theories of temporal reasoning and has thus helped set the agenda for future research.

5.4 Statistical and Probabilistic Theories

While probabilistic and statistical research are an established part of AI research, relatively little work has been done on probabilistic and statistical approaches to temporal reasoning. Two notable works in this area are the probabilistic approach of Pearl (1988, Section 10.4) and the statistical approach of Tennenberg (1991). Both theories are especially formulated to solve the Yale Shooting Problem. Pearl starts off with the YSP in a default logic; there are persistence rules such as $Loaded(t0) \Rightarrow Loaded(t1)$ and $Alive(t0) \Rightarrow Alive(t1)$, and causal rules such as $Alive(t1) \wedge Shoot(t1) \wedge Loaded(t1) \Rightarrow \neg\, Alive(t2)$. These rules are given probabilistic interpretations. Thus, we have:

$P(Loaded(t1) \mid Loaded(t0)) = \text{High} = 1 - \varepsilon$

$P(Alive(t2) \mid Alive(t1)) = \text{High}$

$P(Alive(t2) \mid Alive(t1), Shoot(t1), Load(t1)) = \text{Low}$

Given basic rules of probability theory and some common-sense assumptions on how these probabilities interact, the theory predicts that $P(Alive(t2) \mid Load(t0), Alive(t0), Shoot(t1)) = \text{Low}$ – that is, the theory predicts that Fred will be dead after the shooting.

Tennenberg's theory is based on Kyburg's (1974,1983) theory of statistical epistemology, which argues that agents possess statistical knowledge in their belief bases. For example, [10] an agent might have the statement $P(\textit{mug-in-lounge}(t1) \mid \textit{mug-in-lounge}(t2)) \in (0.9, 0.95)$ in his belief base. An agent reasoning about the Yale Shooting Problem might have statements like:

$P(Loaded(t1) \mid Loaded\ (t0)) \in (0.9, 0.95)$

$P(Alive(t1) \mid Alive(t0)) \in (0.99, 1.0)$

$P(Alive(t1) \mid Shoot(t0) \wedge Alive(t0) \wedge (Loaded(t0)) \in (0.1, 0.3)$

Note that the last two statements are an example of *competing reference classes.* In general, the most specific reference class applicable is used. Tennenberg then suggests ways of extending this approach so that it can handle the Yale Shooting Problem.

The clear advantage of statistical and probabilistic approaches over nonmonotonic approaches is that they are strongly grounded in classical mathematics. But such approaches fall far short of satisfying the list of requirements outlined in Section 4. Tennenberg's approach is highly unintuitive; it is difficult to believe that people assign numerical probabilities to statements. People typically reason with statements like, "The mug will probably be in the lounge in an hour" or "People usually stay alive from one moment to the next." They do not reason with statements such as, "The probability that my mug will be in the lounge in the next hour is between 90 and 95 percent."

Moreover, compiling such a set of statistics seems very difficult. Tennenberg suggests that we learn them informally from observation. Clearly, however, there is much more than observation going on. The ability to reason inductively and analogically is also required. If we know that in 100 previous instances a mug stayed in the room where it was placed, I can reason by induction that when I now place my mug in the lounge, it will remain there. I can presumably reason by analogy and infer that this will also apply to a bowl that is placed in the room. However, I cannot make this same inference about a cat that is placed in the room. [11] Tennenberg's approach also suffers from vagueness; he does not work out

[10] I am using the notation of probabilistic logic here to simplify the exposition.

[11] Interestingly, Tennenberg noted, while discussing frame axioms, that one can say with much more

any of his problems in detailed, concrete terms.

Pearl's approach does not suffer from these defects. The probabilities are derived from the default rules of the theory, so there is no more difficulty in encoding the probabilities than there is in writing down the default theory. Moreover, he assigns vague probabilities like High or Low: intuitively, just the sort that we expect humans to be comfortable reasoning with. The Yale Shooting Problem, at least, is worked out in detail, although none of the other problems in the YSP family are. (Pearl himself notes that his causal rules entail some form of forward reasoning, so it would certainly be wise to ensure that his solution also works for backward temporal projection and explanation.)

Nonetheless, both Tennenberg's and Pearl's approaches fail as solutions to the frame problem in one important respect: they both rely heavily on frame axioms. In each case, a full set of persistence rules is needed. Thus, their approaches do not solve the frame problem at all. [12]

6 Conclusion

Previous solutions to the frame problem have been flawed, mostly because they have attacked too specific or too general a version of the problem. The single largest contributing factor to the misdirected research has been the centrality of the situation calculus to the research on the frame problem. The lion's share of solutions has directly relied on the very strong assumptions of the situation calculus – most notably, the assumptions that concurrent actions cannot occur or that all actions are known – that are out of place in common-sense reasoning. Such solutions are brittle; they collapse as soon as concurrent actions are permitted or unknown actions are possible. This observation underscores the danger of relying too heavily on toy problems in one's research.

Another large segment of research, most notably the monotonic solutions, MAT, and the

confidence that a pair of shorts will remain in some location than that a $100 bill will remain in some location. But he does not seem to realize that a similar problem applies to learning statistical knowledge through observation.

[12] Tennenberg argues that his statistical rules are not as bad as frame axioms because one can make up rules that apply to classes of objects, thus reducing the number of frame axioms. Of course, the same argument can be made for standard frame axioms.

statistical and probabilistic approaches, fail to completely solve the frame problem because they lost sight of what the frame problem is all about – that is, getting rid of the frame axioms. This observation underscores the necessity of clearly defining a problem and stating what the necessary criteria for a solution are.

Despite the fact that solutions to the frame problem have all failed in at least one respect, there exist solutions that are at least partly successful. Very often, these solutions are so brittle that they cannot be extended. However, there is some potential for integrating various solutions. In particular, it is possible to augment MAT – a theory of temporal reasoning that did not solve the frame problem because it relied on explicit frame axioms – with the idea behind Lifschitz's theory of causal reasoning. The resulting integrated theory may very well offer a solution to the frame problem.

More important than the question of whether or not one particular theory is a solution to the frame problem is the lesson to be learned: that new research must build upon previous research, and that integration of existing theories is important. Too little of that goes on in AI logicist research today. The trend is to poke holes in someone else's theory and to propose a theory that is entirely different, rather than to *fix* the holes *within* the theory. The trend should be to modify and integrate. Lifschitz's (1991) and Sandewall's (1994) recent work are welcome examples of this trend.

These, then, are the morals that our investigation has demonstrated:

1. Clearly define the problem one wishes to solve, and write down the set of criteria for acceptance.

2. Resist the temptation to focus on toy problems.

3. Build upon previous research.

These lessons apply not only to the frame problem, but to all of temporal reasoning, and not only to temporal reasoning, but to the entire endeavor of logicist AI. If we have spent the past twenty-six years learning these lessons, they have been twenty-six years well spent.

Acknowledgement: I am grateful to Ernie Davis for his detailed comments and suggestions on an earlier draft of this chapter.

REFERENCES

Baker, A. (1991). Nonmonotonic reasoning in the framework of the situation calculus. *Artificial Intelligence, 49*, 5-24

Baker, A. (1989). A simple solution to the Yale Shooting Problem. In R. Brachman, H. Levesque, and R. Reiter (Eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the 1rst International Conference*, 11-20. San Mateo, CA: Morgan Kaufmann

Cobham, A. (1964). The intrinsic computational difficultly of functions. *Proceedings of the 1964 Congress for Logic, Mathematics and the Philosophy of Science*, 24-30. New York: North-Holland

Cook, S.A. (1971). The complexity of theorem-proving procedures. *Proceedings of the Third Annual ACM Symposium on the Theory of Computing*, 151-158. New York: ACM.

Davis, E. (1990). *Representations of commonsense knowledge.* San Mateo, CA: Morgan Kaufmann.

Edmonds, J. (1965). Paths, trees, and flowers. *Canadian Journal of Mathematics, 17*(3), 449-467.

Elkan, C. (1992). Reasoning about action in first-order logic. In *Proceedings of the Ninth Biennial Conference of the Canadian Society for Computational Studies of Intelligence* (pp. 221-227). San Mateo, CA: Morgan Kaufmann.

Fetzer, J.H. (1991). The frame problem: Artificial Intelligence meets David Hume. In K.M. Ford and P.J. Hayes (Eds.), *Reasoning agents in a dynamic world: The frame problem*, 55-69. Greenwich, CT: JAI Press.

Fikes, R. and Nilsson, N. (1971). STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence, 3,* 189-208.

Finger, J. J. (1988). *Exploiting constraints in design synthesis.* Technical Report # STAN-CS-88-1204, Stanford University.

Gelfond, M., Lifschitz, V. and Rabinov, A. (1991). What are the limitations of the situation calculus? *Working Papers, AAAI 1991 Symposium on Logical Formalizations of Commonsense Reasoning.* Also in R. Boyer (Ed.), *Essays in honor of Woody Bledsoe*, Dordrecht: Kluwer Academics

Ginsberg, M. and Smith, D. (1988). Reasoning about Action I: A possible worlds approach. *Artificial Intelligence, 35,* 165-195.

Haas, A. (1987). The case for domain-specific frame axioms. In F.M. Brown (Ed.), *The frame problem in artificial intelligence: Proceedings of the 1987 workshop*, 343-348. San Mateo, CA: Morgan Kaufmann.

Hanks, S. and McDermott, D. (1986). Default reasoning, nonmonotonic logic, and the frame problem. *Proceedings of AAAI 1986,* 328-333.

Haugh, B. (1987). Simple causal minimizations for temporal persistence and projection. *Proceedings of AAAI 1987,* 218-223.

Hayes, P. (1977). In defence of logic. *Proceedings of IJCAI 1977,* 559-565.

Hempel, C. (1945). Studies in the logic of confirmation. *Mind, 54,* 1-28,97-121.

Kartha, G.N. (1993). Soundness and completeness theorems for three formalizations of action. *Proceedings of IJCAI 1993,* 724-729.

Kautz, H. (1986). The logic of persistence. *Proceedings of AAAI 1986,* 401-405.

Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy, 72,* 690-716.

Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

Kyburg, H. E., Jr. (1974). *The Logical Foundations of Statistical Inference.*, Dordrecht, Holland: Reidel.

Kyburg, H. E., Jr. (1993). The reference class. *Philosophy of Science, 50,* 374-397.

Lifschitz, V. (1986). Pointwise circumscription: Preliminary report. *Proceedings of AAAI 1986,* 406-410.

Lifschitz, V. (1987). Formal theories of action. *Proceedings of IJCAI 1987,* 966-972.

Lifschitz, V. (1988). Benchmark problems for formal non-monotonic reasoning, version 2.0. In M. Reinfrank, J. de Kleer, M. Ginsberg, and E. Sandewall (Eds.), *Non-monotonic reasoning: Proceedings of the 2nd international workshop, Lecture Notes in AI # 346,* 202-219. Springer-Verlag.

Lifschitz, V. (1991). Toward a metatheory of action. In J. Allen, R. Fikes, and E. Sandewall (Eds.), *Principles of knowledge representation and reasoning: Proceedings of the 2nd international conference*, 376-386. San Mateo, CA: Morgan Kaufmann.

Lifschitz, V. and Rabinov, A. (1989). Miracles in formal theories of action. *AIJ Research Note.*

Lin, F. and Shoham, Y. (1991). Provably correct theories of action: Preliminary report. *Proceedings of AAAI 1991,* 349-354.

Lin, F. and Shoham, Y. (1992). Concurrent actions in the situation calculus. *Proceedings of AAAI 1992,* 590-595.

Martin, R. L. (Ed.). (1984). *Recent essays on truth and the liar paradox.* Oxford: Oxford University Press.

McCarthy, J. (1958). Programs with common sense. *Proceedings of the Symposium on the Mechanization of Thought Processes, National Physical Laboratory.* (Also reprinted in M. L. Minsky (Ed.), *Semantic information processing.* Cambridge, MA: MIT Press, 1968, 403-409.)

McCarthy, J. (1980). Circumscription – a form of non-monotonic reasoning. *Artificial Intelligence, 13,* 27-39.

McCarthy, J. (1986). Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence, 28,* 86-116.

McCarthy, J. and Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer (Ed.), *Machine Intelligence* (Vol. 4), 463-502.

McDermott, D. (1978). Tarskian semantics, or no notation without denotation! *Cognitive Science, 2,* 277-282.

McDermott, D. (1982). A temporal logic for reasoning about processes and plans. *Cognitive Science, 6,* 101-155.

McDermott, D. (1987). A critique of pure reason. *Computational Intelligence, 3,* 151-160.

Morgenstern, L. and Stein, L. A. (1989). Why things go wrong: A formal theory of causal reasoning. *Proceedings of AAAI 1988,* 518-523.

Newell, A. (1981). The knowledge level. *AI Magazine, 2,* 1-20.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems.* San Mateo, CA: Morgan

Kaufmann.

Pednault, E. (1991). Generalizing nonlinear plans to handle complex goals and actions with context-dependent effects. *Proceedings of IJCAI 1991,* 240-245.

Penberthy, J.S. (1993). *Planning with continuous change.* Technical Report # 93-12-01, University of Washington, Dept. of Computer Science and Engineering.

Reiter, R. (1991). The frame problem in the situation calculus: A simple solution (sometimes) and a completeness result for goal regression. In V. Lifschitz, (Ed.) *Artificial Intelligence and mathematical theory of computation: Papers in honor of John McCarthy*, 359-380. San Diego: Academic Press.

Reiter, R. and Criscuolo, G. (1981). On interacting defaults. *Proceedings of IJCAI 1981,* 270-276.

Sandewall, E. (1994). *Features and fluents: The representation of knowledge about dynamical systems, vol. I.* Oxford: Oxford University Press.

Schubert, L.K. (1990). Monotonic solution of the frame problem in the situation calculus: An efficient method for worlds with fully specified actions. In H.E. Kyburg, R.P. Loui, and G.N. Carlson (Eds.), *Knowledge representation and defeasible reasoning,* 23-67. Boston: Kluwer Academic Press.

Shoham, Y. (1988). *Reasoning about change: Time and causation from the standpoint of artificial intelligence.* Cambridge, MA: MIT Press.

Stein, L. A. and Morgenstern, L. (1994). Motivated action theory. *Artificial Intelligence, 71,* 1-42.

Tennenberg, J. (1991). Abandoning the completeness assumption: A statistical approach to the frame problem. In K.M. Ford and P.J. Hayes (Eds.), *Reasoning agents in a dynamic world: The frame problem,* 231-257. Greenwich, CT: JAI Press.

Wiles, A. (1995). Modular elliptic curves and Fermat's Last Theorem. *Annals of Mathematics, 141, 3,* 443-572.