AI Magazine Volume 15 Number 1 (1994) (© AAAI)



# Book Review

## Navigating through the Frame Problem

*Selmer Bringsjord and Chris Welty*

■ *Reasoning Agents in a Dynamic World: The Frame Problem,* Ken Ford and Pat Hayes, eds., JAI Press, Greenwich, Connecticut, 1991, 289 pp., $68.50, ISBN 1-55938-082-9.

The frame problem is over two decades old and is still going strong. What accounts for its longevity? Ken Ford and Pat Hayes give the answer in their excellent *Reasoning Agents in a Dynamic World:* The frame problem poses difficulties for AI at several levels simultaneously. The first difficulty is purely definitional: What is the frame problem? Although many thinkers claim to have a firm, intuitive grasp of the problem, when it comes time to deliver specifics, intuitions turn vaporous. As a result, papers written on the subject, whether they propose a solution or propose that there simply isn't a solution, are vulnerable to the common refrain, Yes, but that's not the frame problem.

However, suppose that the frame problem can be defined. Then the next difficulty arises; that is, is a solution to the frame problem required for AI to succeed? Given that there are a goodly number of seemingly intelligent systems doing some pretty clever things without confronting the frame problem, it's tempting to answer this question in the negative. There seems to be a general consensus, however, both in and out of *Reasoning Agents in a Dynamic World,* that ambitious AI can succeed only if the frame problem can be solved. Usually, this belief is expressed by couching the frame problem in terms of a generally intelligent robot. The idea is that although an AI system without the frame problem might, say, read an echocardiogram and diagnose a heart defect, a really smart autonomous robot will arrive only if, like us humans, it can handle the frame problem.

Suppose that a solution to the frame problem is absolutely essential for R2D2's advent (see Dennett [1984]). We're still confronted by a difficult question: Is there a solution to it? If not, then R2D2 might forever be but a creature of fiction. If, however, the frame problem is solvable, we must confront yet another question: Is there a general solution to the frame problem, or is the best that can be mustered a so-called domain-dependent solution?

*The highlight … is an entertaining go-round between two pugilists trading blows in civil but gloves-off style, reminiscent of a net discussion.*

*Reasoning Agents in a Dynamic World* is a collection of essays on the frame problem that arose from the First International Workshop on Human and Machine Cognition held in May 1989. As a collection, it's complete in that it covers the high-level architectonics of the frame problem we just described. There are a lot of forks to ponder in reading the book—whether the frame problem is definable, solvable, a general solution or a domain-dependent one, and so on. We took the approach that because any reading is colored by the reader's prior position, a single review by two people with seemingly orthogonal views on the frame problem would yield an interesting route through the book.

The highlight of the book is the clash between Jim Fetzer and Pat Hayes (Fetzer's "The Frame Problem: AI Meets David Hume" comes first, then Hayes's attack on this chapter, then Fetzer's rebuttal). This set of chapters is an entertaining go-round between two pugilists trading blows in civil but gloves-off style, reminiscent of a net discussion. This point-counterpoint serves as a starting place from which to systematically explore the rest of the book.

Fetzer holds that the frame problem is a special case of Hume's problem of induction, which involves "justifying some inferences about the future as opposed to others" (p. 55), a phrase that, it must be conceded, certainly sounds like the frame problem. Fetzer then proceeds to claim that the problem of induction is solvable only if relevant causal laws are known, where causal laws are conjunctions of factors affecting the outcome. For example, the statement, If one strikes a match, and the match is dry, is of correct chemical composition,..., then the match will ignite, would be a causal law. These laws, he claims, must have maximally true antecedents, that is, "every factor...whose presence or absence makes a difference to the occurrence of that outcome has to be taken into account" (p. 57). The idea here is that our law about match ignition would be counterexampled if we deleted, say, the precondition that the match is dry—because we could then have a case wherein the

antecedent is true, but the consequent isn't.

Hayes's initial reaction, an instance of the refrain mentioned previously, is that the frame problem is not a special case of the induction problem. He gives his own definition of frame problem:

> But the frame problem is not one of justifying certain inferences, but of formulating a succinct and usable way of *expressing* them.... The problem is not an epistemic one of "ascertaining" anything—it is a representational one of how to express the information we want our robots to use. And we have some ideas about how to say which things change. The point of the frame problem is that in the sort of environment we are trying to describe, almost everything *doesn't* change during almost every temporal interval. Surely there must be some compact and principled way of...saying...that things are just normally quiet and uninteresting. This is the frame problem (p. 72).

Is Hayes's definition of *frame problem* definitive? Fetzer doesn't think so. He points out (pp. 78–79) that the frame problem in the literature is maddeningly protean. Here, against Hayes, he's certainly right. Those inclined to side with Hayes might want to consider a Fetzer-supporting string of proposed definitions to be found in Brown (1987) and Pylyshyn (1987). They might also want to consider the Ford and Hayes volume itself, which supports Fetzer's pessimism: Lynn Andrea Stein (p. 219) quips, "a definition of the 'frame problem' is harder to come by than the Holy Grail," and although no one has yet perished in pursuit of a definition, neither has anyone defined it to the satisfaction of all. In fact, by our count, at least 16 definitions of the frame problem are ventured in *Reasoning Agents in a Dynamic World*. (The editors themselves give it a valiant go in their well-written introduction. We recommend that those inclined to study these definitions start with Don Perlis's interesting "Intentionality and Defaults," which nicely complements Elgot-Drapkin, Miller, and Perlis [1987].) It would be overly pessimistic to declare all 16 distinct; however, it's safe to say that there are no clear grounds for holding that the 16—or even some proper subset thereof—phrase the problem as a purely representational one.

There are specific reasons, however, for denying Hayes's claim that the frame problem is solely representational: Suppose that *Q* represents—compactly, elegantly, and rigorously—the notion that, as Hayes puts it, things are just normally quiet and uninteresting. (Frank Jackson's ingenious modal logic *Z*, detailed in his "The Modal Quantificational Logic Z Applied to the Frame Problem," pp. 1–42, which supports most nonmonotonic reasoning theories, is an instance of *Q*.) Now suppose that some robot's beliefs at $t1$ are composed of formulae *D*. Also, suppose that at $t2$, our robot performs some action *a*. It would be exceedingly nice if, at $t3$, the robot's belief set reflects the updated situation resulting from *a*. Actually, that's putting it mildly. It's more than fair to say that arriving at a new belief set for $t3$ very nearly constitutes the frame problem itself (as a number of contributors to this volume agree, including Jackson himself [p. 1]). Here's the key question, though: Will our robot, outfitted with the Hayesian *Q*, be able to propagate *a* through *D* to reach its new belief set? The answer, at least in general terms, is obvious: It depends—because *Q* in and of itself doesn't carry the day. The robot must be able to *use* *Q* to arrive at its new belief set.

*What other blows does Hayes throw against Fetzer?*

It would seem, then, that although the frame problem surely has a strong representational side that would make Hayes happy, it also has a Fetzerian epistemic side on which the representational side must be based. The frame problem is not purely representational, and neither is it purely epistemic. This brew implies, against Fetzer, that the frame problem is not simply a special case of the induction problem.

However, Fetzer needn't stick to his view that the frame problem is a special case of the induction problem. In fact, his writing suggests that he would be comfortable retreating to the view that the frame problem is solvable only if the induction problem is. The rationale for this condition, given the previous discussion, seems, in general terms, a straightforward application of conditional proof: Suppose our hypothetical robot can solve the problem of propagating action *a* to $t3$. Then it can use *Q* to justify some inferences about the future as opposed to others; that is, it can use *Q* to solve the induction problem.

What other blows does Hayes throw against Fetzer? Hayes says (p. 73) that causal laws can't be known if their antecedents are, to use Fetzer's term, maximally specific because the antecedents would be too large to be commonsensical. (Thus, in the match law case, the claim would be that there are just too many tangential things in the law's antecedent for any human to grasp and deploy this law.) This attack sounds promising, but it gives rise to a rather interesting conundrum for Hayes: Hayes believes that the frame problem is solvable; so, let's grant him that. We've seen that it's plausible that if the frame problem is solvable, then the induction problem is too. However, if Hayes is correct that causal laws can't be known, it follows that the induction problem is unsolvable. It then follows, by modus tollens, that the frame problem itself is unsolvable! Thus, Hayes finds himself in the unenviable position of having to swallow something that's not particularly savory: a contradiction.

Hayes might have a clever escape from this riddle up his sleeve. Assume, he says, that we have command over all the causal laws for physics, biology, chemistry, and so on. "How would these help us predict the future? We would still need to know which aspects of the world in which we live the laws applied, and which aspects were irrelevant" (p. 76).

This objection is ambiguous because of at least two interpretations. On the one hand, Hayes could be saying that command over the causal laws is insufficient for solving the induction problem (and insufficient for solving the frame problem). On the other, he could be saying something much stronger: that having command over all the causal laws is thoroughly irrelevant to the frame problem. If Hayes intends the former, he's in trouble because Fetzer would no doubt be the first to agree that knowledge of relevant causal laws is insufficient for solving the problem of induction. One would obviously need, in addition to such knowledge, the ability to carry out deduction, to remember, to calculate, to observe, to read, and so on. If Hayes intends the latter, he's still in trouble because having command over the relevant information isn't irrelevant to the frame problem. After all, if R2D2 knows that part of its knowledge about causal laws is relevant to tackling an instance of the frame problem (which, with the interpretation we're considering, is what Hayes says is tantamount to solving the frame problem), it follows that R2D2 knows certain causal laws, which is the knowledge Fetzer thinks important.

Thus, we are forced into a third reading of Hayes's complaint, one that doesn't offer an escape from the aforementioned conundrum and produces considerable irony: To be plausible, Hayes's objection must amount to the claim that if we know all causal laws, we must still solve a remaining part of the frame problem to solve the induction problem. If such is the case, and others (for example, van Brakel [1992]) seem to think it is, we're still left with the contrapositive of Fetzer's conditional: If the induction problem is unsolvable (as Hayes would have us believe), so is the frame problem. Neither Fetzer nor Hayes seems to think the induction problem (or the frame problem) is unsolvable, just that the other doesn't know how to solve it. Their positions, however, seem to point strongly to the conclusion that there is no solution to either the induction or the frame problem!

Now, we're not saying that every open-minded, rational reader of Ford and Hayes's book ought to come to the conclusion that the frame problem is unsolvable. However, we are saying that Fetzer's reasoning is powerful and thought provoking. Taken in conjunction with other issues raised by this book, it should at least raise one's eyebrows to the possibility that the frame problem might indeed have no solution.

What other issues are raised in this book? Well, the volume is reminiscent of situations in microphysics where the attempt to measure a system renders it unmeasurable. What's the analog to the microworld in *Reasoning Agents in a Dynamic World*? It's that the very act of attempting to pin down and solve the frame problem spins off a bewildering array of variants on the problem, each itself quite a challenge. The best example of this Heisenbergian phenomenon is Leora Morgenstern's "Knowledge and the Frame Problem." She considers variants on the frame problem that arise when a logic of knowledge is combined with a logic of action that allows for multiple agents. She calls one of these variants the *third-agent frame problem;* it runs roughly as follows: Suppose an agent $S$ is constructing a plan to bring about $p$, which requires the participation of another agent $S'$. How can $S$ be sure that $S'$ knows enough to play his or her role in the plan in question?

Now, our claim is not that the three-frame problem isn't solved by Morgenstern. Rather, our claim is that Morgenstern provides absolutely no reason to think that her general approach to the frame problem won't continue to limitlessly spin off virulent variants of it. In fact, a pervasive brittleness to Morgenstern's approach suggests that it will need to be, to put it mildly, fine tuned. For example, nearly all the psychological axioms in her theory can be counterexampled with real-life scenarios.[1] Moreover, there is no reason to think that the trend of combining logics to increase expressivity will stop any time soon. We seem to need, for example, for a robot capable of the complicated behavior that Morgenstern envisions, logics not only of knowledge (and belief) and action but, say, logics that cover more sophisticated conditionals than her system allows (see, for example, Nute [1984]) and also, perhaps, logics that cover deontic notions (see Chellas [1980] for the simplest deontic logics). If such formalisms are added to the brew, it's seems safe to conclude that more nasty variants will be spawned. Where does it all stop? Does it stop?

Suppose it doesn't stop. Then, like one of us (Bringsjord), you might despair of ever seeing robots that are persons (and you might have reasons for such despair that have nothing to do with the frame problem [Bringsjord 1992]). However, suppose, like the other of us (Welty) and like most others in AI (for example, Terry Nutter, who in her suggestive "Focus of Attention, Context, and the Frame Problem," starts by granting the frame problem's unsolvability), you don't intend to let these theoretical worries stop you from building systems that are precursors to a generally intelligent robot. What should you do? You probably ought to try to capitalize on domain-specific knowledge. A good place to turn in Ford and Hayes's book is Jay Weber's "The Myth of Domain-Independent Persistence," which presents (pp. 265–267) what students in introductory AI classes press against their instructors regularly: Even the famous Yale turkey shoot (set out in Hanks and McDermott [1986] and elegantly discussed in Loui [1989]), taken by many to be the benchmark encapsulation of the frame problem, is solvable if you include the sort of domain-dependent knowledge that we have about such scenarios. A

related approach is based on the observation that a proposed solution to the frame problem needn't be correct all the time. After all, perhaps humans have no perfect solution, simply techniques that work often enough to allow homo sapiens to survive. This statistical approach to the frame problem and, specifically, to the Yale turkey shoot, is discussed in Josh Tenenberg's "Abandoning the Completeness Assumptions: A Statistical Approach to the Frame Problem." This approach seems to imply that theoreticians are too busy mourning the death of proposed general solutions to the frame problem at the hands of counterexamples to see that these counterexamples are unlikely.

Although there seem to be acute problems facing system designers who seek to dodge the frame problem in these ways,[2] perhaps one of the dodges will work—there's certainly some hope. We encourage you to read Ford and Hayes's excellent volume and decide for yourself.[3]

## Notes

1. For example, weakness of the will, a condition that unfortunately afflicts a good many of us, would counterexample Axiom Goals2: If *b* has the goal of performing a certain action and can perform that action, he will perform that action (p. 164). For example, Jones has the goal of talking to his mother-in-law in a normal voice, and he can do so (his vocal chords are fine, and so on), but, alas, Jones loses his temper once again and yells at her.

2. For example, the domain-dependent knowledge is usually defeasible, as in this example of the turkey shoot: An agent must be grasping a gun to unload it or shoot it, which is potentially defeated by the fact that guns can be fired by remote control. In addition, the statistical approach requires agents to have extensive knowledge of probabilities—knowledge so extensive that it can rival that required in frame-axiom approaches.

3. Thanks are owed to the computer science graduate students whose reaction to the Ford and Hayes volume in a spring 1993 seminar helped one of us (Bringsjord) evaluate this book. Thanks also to Marie Meteer, who co-taught the seminar with Bringsjord.

## References

Bringsjord, S. 1992. *What Robots Can and Can't Be*. Dordrecht, The Netherlands: Kluwer.

Brown, F. M., ed. 1987. *The Frame Problem in Artificial Intelligence: Proceedings of the 1987 Workshop*. San Mateo, Calif.: Morgan Kaufmann.

Chellas, B. 1980. *Modal Logic: An Introduction*. Cambridge, U.K.: Cambridge University Press.

Dennett, D. 1984. Cognitive Wheels: The Frame Problem of AI. In *Minds, Machines, and Evolution,* ed. C. Hookway, 129–151. Cambridge, U.K.: Cambridge University Press.

Elgot-Drapkin, J.; Miller, M.; and Perlis, D. 1987. The Two Frame Problems. In *The Frame Problem in Artificial Intelligence: Proceedings of the 1987 Workshop,* ed. F. Brown, 23–28. San Mateo, Calif.: Morgan Kaufmann.

Hanks, S., and McDermott, D. V. 1986. Default Reasoning, Nonmonotonic Logics, and the Frame Problem. In Proceedings of the Fifth National Conference on Artificial Intelligence, 328–333. Menlo Park, Calif.: American Association for Artificial Intelligence.

Loui, R. 1989. Back to the Scene of the Crime, or Who Survived Yale Shooting? Presented at the International Workshop on Human and Machine Cognition, Pensacola, Florida, 11 May.

Nute, D. 1984. Conditional Logic. In *Handbook of Philosophical Logic, Vol. 2,* eds. D. Gabbay and F. Guenthner, 387–440. Dordrecht, The Netherlands: D. Reidel.

Pylyshyn, Z. W., ed. 1987. *The Robot's Dilemma*. Norwood N.J.: Ablex.

van Brakel, J. 1992. "The Complete Description of the Frame Problem." *Psycoloquy* 92.3.60 (frame-problem.2).

**Selmer Bringsjord** is the author of the monograph *What Robots Can and Can't Be* (Kluwer, 1992). His current research interests include AI and infinitary logics and the notion of point of view in the context of an attempt at Rensselaer Polytechnic Institute to design computerized story-generation systems.

**Chris Welty** works in the Rensselaer Polytechnic Institute Computer Science Department, where he is also completing work on his Ph.D. in AI and software engineering. He is a member of the American Association for Artificial Intelligence and serves on the editorial board of the *SIGART Bulletin* for the Association for Computing Machinery.