

The Frame Problem Blues. Once more, with feeling.

Zenon Pylyshyn
Center for Cognitive Science
Rutgers University

For many of the authors in this volume, this is the second attempt to explore what McCarthy and Hayes (1969) first called the “Frame Problem”. Since the first compendium (Pylyshyn, 1987), nicely summarized here by Ronald Loui, there have been several conferences and books on the topic. Their goals range from providing a clarification of the problem by breaking it down into subproblems (and sometimes declaring the hard subproblems to not be the *real* Frame Problem), to providing formal “solutions” to certain aspects of the problem. But more often the message has been that the problem is not solvable except in a piecemeal way in special circumstances by some sort of heuristic approximations. It has sometimes also been said that solving the Frame Problem is not only an unachievable goal, but it is also an unnecessary one since *humans* do not solve it either; we simply get along as best we can and deal with the problem of planning in ways that, to use Dennett’s phrase, is “good enough for government work”.

Whatever approach one takes it seems that the general issue will not (or does not) go away. None of the proposals settle the main questions: What is the genus to which the species “Frame Problem” belongs? Do we have any idea how to deal with it even in the approximate sense that humans may deal with it? Can we deal with it in an approximate way that does not break down and lead to obvious and egregious errors on simple counterexamples. Would it help if we developed some novel alternatives for the form, content or medium of the representation we use, or of the reasoning systems? Can we get along without solving the Frame Problem by some clever finesse in, say, the way we (or the robots) formulate the planning or the reasoning problem? Or perhaps, as Rod Brooks has argued (Brooks, 1991), organisms merely have a hierarchy of reactive mechanisms with no representations at all. If that is the case then no frame problem arises since no abstract reasoning about the future occurs at all.

The first question — what general problem is the Frame Problem an instance of? — is not a question of getting the definition right. Leola Morgenstern’s chapter provides a clear sketch of the geography of the problem and discusses the pitfalls of taking a too broad or too narrow definition of the Frame problem. Of course we could (and many do) define the Frame Problem to be some special technical subproblem faced by robots which plan their actions from representations of the world formulated in some formal logical notation. Yet if it turns out that the major problem in planning and reasoning about the consequences of actions is, say, exactly

the problem of making inductive inferences, or the problem of deciding what is relevant in solving problems or predicting the future course of some system, then we will have discovered that the problem that arose in robotic planning is a special case of a more general problem. If this is true then we can be sure that no partitioning of the problem or defining it differently will deal with the basic source of difficulty. It would be like discovering that, say, the constraint satisfaction problem is computationally equivalent to the travelling salesman problem. It tells us that there is no algorithm that can solve it in general in polynomial time. If we fail to recognize this then what can, and does, happen is that we find ourselves pursuing a shell game whereby we deal with a portion of the problem by hiding the hard part in another place. If you squeeze in one place the hard problem just balloons out in another.

There is a shaggy-dog story, which I first heard from Hilary Putnam, about an engineer who claimed to have invented a perpetual motion machine. His boasting finally became unbearable so his friends agreed to visit his basement and see the machine he had invented. The engineer proudly escorted them to his basement laboratory which was filled with all kinds of marvelous glistening equipment; precision-made gears, levers, pulleys, and even solid state semi-conducting devices with multicolored lights. His audience marveled in silence at the array of breathtaking and colorful equipment. Finally one of them got up the nerve to ask the question that was on everyone's lips: Since this was supposed to be a perpetual motion machine, why was nothing moving? "Oh that", replied the engineer, with undiminished enthusiasm, "Well the machine is all finished except for one minor little piece that is still on order and is expected any day now. It's just a small ratchet that fits in here and goes back and forth forever."

So it is with solutions to the Frame Problem. They usually end up with the critical element being on order. As Janlert puts it in his chapter, there are a series of closely interconnected problems involved here and "There would be little point in 'solving' the Frame Problem if that meant 'unsolving' some other problem". Yet more often than not this is what happens, as Fodor argued in characteristic Fodorian style in the 1987 book. The magic missing element can be a new ontology or categorization of the world's properties, as Hayes once argued (Hayes, 1985). Or it can be a missing theory of causal reasoning — even approximate causal reasoning — as Morgenstern and Glymour believe. Or it may be that we need to get away from logical calculi entirely and use some analog medium of representation for reasoning about change — a medium whose properties ensure that the representation remains faithful to the represented world through a natural harmony of causal correspondence. This alternative has been particularly popular among psychologists like Kosslyn (19xx) who postulate a special analog medium for imagery, though philosophers like Haugeland and Dreyfus advocated it in the earlier collection and both Janlert and Dennett flirt with the notion in their current chapters. Of course another possibility is that the solution to the Frame Problem may require a hybrid of logic and analogs, as some people have suggested.

The latter has always been a very tempting idea. The proposal is to leave to some form of combinatorial symbolic encoding (i.e. a logical calculus) the more abstract elements of reasoning

as well as the general principles by which certain domains are handed off to special purpose analog mechanisms. These analog mechanisms can incorporate in their very structure the principles that would otherwise need to be enshrined in so-called Frame Axioms, principles which tell you what does not change when some things are changed. For example, if the spatial layout is represented in terms of locations in a 2 (or 3) dimensional rigid structure (like a blackboard) then we don't need to reason about all the relative locations of irrelevant objects that are not changed by the movement of some of the objects. Once we locate parts appropriately and specifically state the principles for interactions among objects when causally connected objects are explicitly moved, *all* aspects of the relative locations of objects can be merely read off the representation. The "sleeping dog" strategy (as McDermott called in his 1987 chapter) applies by default because of the physical properties of the rigid analog medium. Moreover when you move something, the relative location of every element will be correct and can be merely read off. Note that this is not the case in a propositional representation. Even if you use the "sleeping dog" strategy and do not alter the locations of objects that are not connected to explicitly moved objects, the relationships among all objects have to be inferred after each move (not to mention after each new piece of information about relative location is added, even if nothing was moved). For example if one moves object P1, then any geometrical relationships into which it enters (and there are an indeterminate number of them) could change. If P1 was collinear with P2 and P3 this may no longer be true, if P1 was inside the convex hull defined by P2, P3, P4, ... Pn this may also no longer be true, and so on without limit.

In an appropriate analogue medium it is straightforward to arrange that only relevant changes occur when some particular object is moved. You don't have to worry about objects that are not causally connected or objects that do not lie on the path of the moving object. In this kind of representation one appears to get the geometrical properties free (and even a few physical properties if such restrictions as non-interposition of points are built in). But there are many obvious reasons why this does not work (see e.g., Pylyshyn, 1972, 1981, 1984) so proponents of such "imagistic" or pictorial representations find themselves postulating an array of ad hoc devices to deal with the overly-rigid aspects of such analogs (Kosslyn et al).

Analog geometric media also allow you to read off too much and with too much precision while not providing the needed flexibility and vagueness such as that which you can get with propositional representations where, for example, disjunction is allowed (e.g. you can represent A or B in propositional form, where A and B are quite different situations, but you cannot do so in analog-pictorial form, at least not without some additional apparatus). You cannot represent indefinite or incomplete geometrical relationships — such as A being "next to" B. In order to do that you not only have to have a representation for this state of affairs, but it must behave correctly when new information is added. For example if you know that A is between B and C and you discover that it happens to be collinear with B and C you must be able to infer that it is on a path from B to C, and so on. While the rigidity and fixed-function of analogs must be somehow relaxed, at the same time we also want to retain their attractive automatic geometry-

maintaining features. This is the Freedom versus Stability dimension that Janlert speaks of — in this case applied to geometrical properties. Not only do you need the features of both, but you also need a way to interface them that itself does not presuppose the solution to the tractability of inference problem — and hence to the Frame Problem. In his chapter, Loui makes a similar point when he locates the complexity (and flexibility) of analogs in the way they are validated or mapped into their semantic domain. Inasmuch as it is the organism (or robot) that must do that mapping this observation also locates the frame problem — and tractability in general — to what I called the interface between the analog and the non-analog reasoner.

This conflict between the desiderata of flexibility and tractability is quite general. In Cognitive Science one wants the most constrained possible structure compatible with the observed behavioral capacity. This provides the most powerful explanation because it has the fewest empirical degrees of freedom: it accounts for the largest range of phenomena with the fewest free parameters. Yet the flexibility is essential: You can't reason — even in a domain as apparently restricted as geometry — with only an analog representation such as a diagram. One obvious way to go is to combine the analog mechanism with a general problem solver that is smart enough to know when to use the analog and when to depart from it (presumably this is what we do when we use diagrams in proving theorems in geometry). Another is to only map a restricted set of properties onto the analogue — such as for example, mapping the topology of the represented world onto the neighborhood properties of cellular automata. In general something of this general hybrid sort MUST be the case since computation itself presupposes an architecture with some fixed structure that carries out the symbolic manipulations. Such fixed causal structure may be viewed as analog (as I do in Pylyshyn, 1984) even though it need not have properties of continuity. But this is a far cry from solving the Frame Problem by using analog representation. Indeed, all the issues associated with the Frame Problem reappear in such hybrid systems simply because they still require a general reasoner to oversee the use of the analogs.

Many writers (in this book this is true of Dennett, Dietrich & Fields, Glymour, Kyberg, Lormand ...) take it for granted that some approximate way of updating a database and maintaining its consistency — using heuristics or some approximate inductive methods — will allow us to deal in practice with the Frame Problem, if not to solve it normatively. But it is not obvious that this will do. Janlert explicitly denies this approach, claiming that the Frame Problem is NOT a problem of how to engage with the world, but a problem of how to maintain a representation in step with a changing world. Whatever the ultimate merits of this argument, one thing appears amply clear from the recent history of the study of general problem solving and reasoning in AI: Quick fixes like Scripts or frames or analogs always leave the main problem somewhere else (“on order” as it were). All these hybrid methods beg the question: How do you know which cheap trick to use on a particular occasion (the *indexing* problem), how do you deal with the problem of mapping a particular instance onto such a trick (the *mapping* problem), how do you deal with instances that deviate from the canonical cases covered by the tricks (the *customization*

problem), and how do you in general interlace the capacity provided by the trick with general intelligence (the *interfacing* problem).

A concrete example of the latter issue arose some years ago when researchers in language parsing attempted to make parsing more efficient by storing phrases (using a so-called “phrasal lexicon”). While this seems like a good idea it turns out that while we save some parsing effort, the problem of how to put the phrasal components together is every bit as difficult as the original problem of parsing the string of words. Of course it turned out that if we approach parsing from the point of view of taking a conventional grammar and parser and compiling in some short cuts for frequently encountered strings of form classes (as was done in “chart parsers”) then there are some efficiencies to be gained. But these are marginal efficiencies within the original problem class, not a radical new way of doing things. This is the most likely eventuality in the use of any of the quick fix heuristics, once we agree to open up the domain to which they are applied. Of course if we keep the domain artificially restricted (which is not a bad idea for practical commercial applications) then we might never encounter the combinatorial problem. But even here, the problem of finding a partition of subdomains so that they have corresponding tractable strategies is a non-trivial one. Indeed, this idea has a long history and was referred to as the “big switch” theory of intelligence (Newell, 19xx). The trouble with that strategy is twofold: How does one decide which position of the Big Switch is the appropriate one for some novel problem, and how to coordinate the community of specialists to produce a result that depends on some mix of their expertise.

One of the things that makes the Frame Problem so deceptive, and why so much disagreement arises concerning whether it needs to be solved or whether it is of concern to anyone but philosophers (a debate that pervades the 1987 compendium and some of this one as well) is the fact that it is easy to lose sight of which shell the problem hides under when one appears to have a proposal for dealing with it (e.g. as in the “sleeping dog” strategy). It is reminiscent of some debates about whether natural language inputs need to be fully parsed. There have been proposals through the years of pragmatically motivated approaches to analyzing sentences for such purposes as database queries, information and document retrieval, and automated translation. Such approaches recommended keyword “grammars”, phrasal grammars (alluded to above), statistical analyses, and other so-called engineering approaches (I recall being told by a prominent machine translation researcher in the 1960s that the reason the problem appears so intractable is that it has been in the hands of linguists and philosophers rather than engineers).

Short cuts for such problems always appear attractive because they can be tailored to deal with subsets of a problem. But no matter how well they deal with subsets and no matter how many such subsets are covered, it always turns out that what has been left out contains the original problem all over again.

=====

The present collection of papers provides an overview of current thinking (I hesitate to say “of the Frame Problem” since much of the discussion concerns the question of what problem this really is) in the light of some of the things that have happened since the 1987 collection. While it does not solve it (though some, like Glymour, believe they have made a start on it through work on automated induction) it does show that recognition of the breadth of the problem is more widespread than it was in 1987 when a lot of time went into arguments about whether one discipline or another knew that it was talking about.