

FRAMED

(presented at MIT, May 1984)

Kent Bach

I once asked an eleven-year old girl called Polly, “what problem is everyone faced with at every waking moment?” She immediately replied, “the problem of what to do next.” Imagine my delight when I ran across the last line of Newell and Simon's Turing Award Lecture (1975), where they say, “the critical question is always: What to do next?” Polly was right: we are always faced with the problem of what to do next, certainly if doing includes mental actions, such as directing one's attention or making a decision. Of course, the problem of what to do next breaks down into an endless succession of little problems. As agents, we tend not to notice these little problems, which are overshadowed by those on which we deliberate. Theorists also focus on deliberate thought and action, pretending that a person is always faced with well-defined alternatives and makes a judgment or a decision based on an explicit evaluation of them. But all this takes time, and at any given moment the problem of what to do next must be solved immediately. Usually we solve it without much trouble, thanks to our extensive knowledge of what to do when. Relevant here is Herbert Simon's (1947) observation that people generally don't optimize but merely “satisfice.” I think that people satisfice not merely because they are satisfied with less than the best, or because the best is too hard to find or not worth the extra bother, but because, owing to the ongoing problem of what to do next, satisficing is necessary.

On a different but related theme, a few years ago I wrote a little paper called “Taking Things for Granted.” It dealt with the notorious Gettier problem in the theory of knowledge. I put up a back-handed defense of Gilbert Harman's solution, which was based on his psychologistic strategy of using intuitions about knowledge “to decide when reasoning has occurred and what reasoning there has been” (1973, 47). Harman's solution was that reasoning can give one knowledge only if it contains no false steps. The trouble, so it seemed to me, was that to make Harman's solution work, too much reasoning had to be ascribed to

the reasoner. Since the reasoning in a Gettier case has to be elaborate enough to contain the requisite false step, parallel reasoning in a normal case leading to genuine knowledge has to contain a counterpart of that step. However, this seemed psychologically unrealistic, in that the required step concerned a proposition which is not explicitly considered but merely taken for granted. So I tried to vindicate Harman's approach by distinguishing between reasoning realizing an inference pattern and merely its instantiating that pattern. Reasoning could instantiate a pattern without realizing it if did not contain elements corresponding to all the steps of that inference pattern. I argued that all this makes sense if our reasoning operates according to what I call the taking-for-granted rule.

(TFG) If it seems to me that p , then infer that p ,
provided no reason to the contrary occurs to me.

I suggested that routine reasoning relies on the TFG and that this reliance is justified insofar as one is able to detect abnormal circumstances. That is, I must be pretty good at knowing when not to infer that things are as they seem in order to be justified, when the situation is normal, in supposing that things are as they seem. If I were insensitive to abnormal situations, I would directly infer that p even when I should not. In following TFG, whenever I directly infer that things are as they seem, i.e. without considering reasons to the contrary, I implicitly rely on my reliability at detecting indications of abnormality.

When I wrote "Taking Things for Granted," I was pretty ignorant of AI. So I was pleased to learn, as many people have pointed out to me, that I was on to what people in AI call the frame problem. Today I would like to relate the TFG rule to the frame problem and to some other ideas current in AI, including frame systems of representation. Since I know of no good reason to regard the processes leading to belief as fundamentally different from those leading to decision or action, I will use examples of both to support a conception of ordinary human inference that differs radically from common philosophical views. This conception seems much more realistic psychologically and, moreover, comports nicely with recent work in AI.

As I understand it, there are really two sides to the frame problem: getting a system to make relevant inferences and keeping it from making irrelevant ones. We are pretty good at that. For example, if you have just dressed for a dinner party and then discover that your car isn't running, you realize that if you take your motorcycle, you'll mess up your tuxedo. So you call a cab. Your beliefs also warrant the inference that you are not wearing a football uniform. Yet this thought, not to mention countless others that are equally warranted, does not occur to you. That is fortunate, since there is no end to the number of irrelevant inferences you could make. Note that not everyone is good at making relevant inferences and at not making irrelevant ones. Certain obsessive-compulsives have an extremely liberal notion of what is worth considering. To them almost anything is relevant. As a result, even the simplest task can take forever. In contrast, impulsive people fail to consider relevant consequences of their thoughts or deeds.

For AI the frame problem is how to get a system to make relevant inferences without also making irrelevant ones. But how do you get a system to do that unless it has some way of identifying what is relevant? Virtually anything that is generally irrelevant to a given task can be relevant to it in some contexts. Moreover, as Marvin Minsky (1981, 125) has pointed out, employing rules of relevance and irrelevance, would result in "annoying statements about ... irrelevancy." Indeed, I should add, most of these statements would not just be annoying but themselves irrelevant (except for statements of irrelevancy that concern ostensibly relevant considerations).

It seems plausible to suppose that people operate with heuristics or rules of thumb, call them what you will, that abbreviate processes of inquiry, inference, and deliberation. Like *ceteris paribus* statements, they are reliable but not foolproof. Fortunately, when things go wrong, backup systems, invoking troubleshooting heuristics, can come to the rescue. How this happens is itself part of the frame problem, for it too involves questions of when to use available information and which inferences to make. Recent work in AI (e.g. Lenat 1982, 1983 and Aikins 1983) suggests that this is not an impossible problem. At any rate, heuristics, like

proverbs, can serve us well even when they conflict with one another. A classic example is the pair, "Look before you leap" and "He who hesitates is lost." Of course they don't really conflict if they are not meant to apply in the same situations. If you're a tournament chess player, as I once was, you know to doublecheck your move before making it. On the other hand, if you're in time pressure, with many moves to make in a few seconds, you'd better make the first move that comes to mind. In those circumstances, he who hesitates is lost by forfeit. However, these are not the only two types of circumstances. Before the onset of time pressure, you're in the position of knowing that the more time you take on your present move, the less time you'll have for your later moves, but you know also that a blunder now would end the game without those later moves ever being played.

This example raises the question of how do you decide whether to act immediately, doing the first thing that comes into your head, or to ponder. But putting the question that way is paradoxical, for it suggests that when you act without hesitating, you do so as the result of deliberation. Besides, it suggests that when you do decide to act immediately rather than to deliberate, you make two separate decisions, as if you first decide to act immediately and then you decide what to do. Surely this picture is all wrong. Rather, you do what occurs to you to do unless it immediately occurs to you not to. In practice, this might involve not the thought not to do it but the thought of an alternative or the thought of a reason against doing it. In any case, the psychological path from thinking of it to doing it is broken by some intervening thought. I am suggesting, then, that at every moment we next do the first thing that occurs to us, provided this thought is not immediately overridden by a further relevant thought. Of course, there are many things that we do as the result as deciding to do them, but most of what we do we do because we don't decide not to (as noted in Bach 1978, most philosophers of action arbitrarily restrict their theorizing to deliberate actions). That is, generally we do something because it occurs to us to do it and it does not occur to us not to. On my word processor I have to punch the EXECUTE key in order to get it to do anything, but I don't have to do any such thing to get myself to do something. For example, I feel cold,

so I get up and close the window. It doesn't occur to me not to. That might have occurred to me, but it doesn't. So I act. The same thing happened to William James when staying in bed one cold morning. Every so often, he thought about getting up. Eventually he did. Evidently, every previous time that he thought about getting up he thought also of a reason not to and didn't. The last time no such thought occurred.

In contrast to immediate actions, what about actions to be performed later? Suppose you are trying to decide what to do some bright Sunday afternoon. How do you know when to stop thinking about it? At any given point, you might be leaning toward one particular option, say playing Frisbee on the bank of the Charles, but you have not yet settled on that option. Nor have you set a deadline yet for your decision (you might, when it begins to appear that your opportunities are slipping away). Indeed, there seems to be no point at which you have to make up your mind. The decisionmaking process is open-ended, in the sense that more can always go into it: invention of new options, deeper search into the consequences of the old options, reevaluation of your assumptions or preferences. Presumably the length and the depth of the decisionmaking process depends on the assumed importance of the decision, together with situational time limits, but even so, there is generally no precise point at which the process must end. Yet it often does end. How?

One suggestion is that there is an ongoing measure of the decision cost, which is compared to the value of the decision itself. When the cost gets too great relative to the value of the decision, which itself might be diminishing, as opportunities are cut off and options are foreclosed, a decision is made: the option highest on the list gets selected. That seems plausible enough, but who knows if it really happens? Here's another suggestion. The decision is made when, after exactly one option is at the top of the list, the decisionmaking process stops. It stops when no further considerations occur to one: no further options, no reconsideration of previously considered options, no reasons for not selecting the currently preferred option. In other words, one decides on the preferred option when nothing against it

occurs to one. At that point, one selects it by default. I suggest that practical reasoning is governed by a counterpart to TFG, which I dub the not-worth-considering rule:

(NWC) If it occurs to me to do A, do A (or, if A is a future action, decide to do A) unless the thought of a reason to the contrary (or of an alternative to A) occurs to me.

Earlier, in discussing immediate actions, I suggested that generally we do something because it occurs to us to do it and it does not occur to us not to. Now consider some cognitive analogs of this process, as in recognition and recall. You see someone and take him to be your dentist. It does not occur to you that he might not be, and so you suppose that he is. Or if someone asks you, "What is the capital of Nova Scotia?" A certain answer occurs to you, no alternative occurs to you, it does not occur to you that this might be wrong, and you answer "Halifax."

These examples of immediate practical and factual reasoning suggest a certain view of ordinary inference that departs radically from traditional conceptions. The closest nontraditional conception I can think of is Gilbert Harman's view of factual reasoning as "inference to the best explanation." The trouble with that is that we are usually satisfied with less than the best: Harman should have said "inference to a plausible explanation." To put it in a way that covers practical reasoning as well, we could call it inference to the first unchallenged alternative. By that I mean the first option that comes to mind without being accompanied or immediately followed by the thought of any reason against it or of any alternative to it. That is, we jump to conclusions except when we look before leap. Similarly with routine practical reasoning. It commonly seems to be a case of inference to the first unchallenged option. That is, we decide to do or, in the case of thinking about what to do next, we just do the first thing that comes to mind without being accompanied or immediately followed by the thought of any reason against it or alternative to it.

In the case of factual reasoning, I am suggesting that we jump to conclusions except when we look before we leap. That's obviously efficient, but how reliable is it? Offhand,

jumping to conclusions seems to gain speed at the risk of error. It looks as though it could get us into lots of trouble. But don't forget, drawing inferences is, as Mill observed, "the only occupation in which the mind never ceases to be engaged." We can't avoid trading off possible error for speed, for there are always more inferences to be made. This means that if we didn't generally jump to conclusions, we wouldn't make most of the inferences we need to make. In any case, it seems that when we jump to conclusions, we are generally right. We are generally right in our snap judgments about the kinds and qualities of things we perceive around us, right in our recollections about prior experiences, right about persons, places, and things we seem to recognize, right about what people mean when they talk to us. Perceptual judgment, recall, recognition, and understanding utterances (see Bach and Harnish 1979) are clear cases of generally reliable jumping to conclusions. Since this is not a monumental coincidence, somehow our inferences must take relevant information into account without getting bogged down in irrelevancies. But how? How do we resolve the tension between efficiency and reliability? It seems that reliability requires justifiably ruling out alternatives to the tempting conclusion. The way this tension is resolved, I suggest, is that alternatives can be effectively and legitimately ruled out without even being considered, at least not consciously. This can occur if our reasoning processes have the following feature: we consider an alternative only when there is special reason to do so. Otherwise, without explicitly thinking that the alternative does not obtain, we reason as if it does not.

Now obviously our reasoning can work like this only if we are equipped somehow to detect the presence of reasons for considering alternatives that we ordinarily take for granted not to obtain. A belief resulting from such a process is justified to the extent that the process not only leads to true beliefs, at least generally, but also guards against forming false beliefs by means of precautionary subroutines that are generally activated when and only when they need to be. Thus I suppose that ordinary abbreviated reasoning operates according to the taking-for-granted rule.

(TFG) If it seems to me that p , then infer that p .

provided no reason to the contrary occurs to me.

TFG licenses me to jump to conclusions if I don't think of a reason not to. This means that the justification of such an inference is conditional on the nonoccurrence of a certain thought. In the case of visual belief, for example, ordinarily I assume that things are as they look, unless it occurs to me that my vision is being affected abnormally, say by bad lighting or devious psychologists. Similarly, in the case of recall, as of somebody's name or the spelling of a certain word, I take for granted that the first thought that comes to mind is the right one-- unless it occurs to me that it might not be, say because some other possibility comes to mind.

Since making inferences according to TFG requires the nonoccurrence of a certain thought, TFG has the remarkable feature that it cannot be explicitly followed. For if TFG occurred to me while I was following it, then I would have to consider whether there are occurring to me any thoughts to the contrary of my prospective conclusion, in which case I would no longer be drawing that conclusion directly. Instead, my reasoning would contain the additional thought that there are no reasons contrary to that conclusion. But that's not the way jumping to conclusions goes, or at least not the way it seems to go. I don't seem to draw my conclusion after noting that no contrary possibility has occurred to me, and if I did reason in that way, undoubtedly plenty of such possibilities would occur to me.

From the programming point of view, what seems not to be going on is that I am operating with a huge number of IF THEN statements followed by an ELSE statement [infer p], where each IF clause contains some adverse possibility which the corresponding THEN clause says how to deal with, when detected. On that model, if no condition specified by an IF clause is satisfied, then the ELSE clause goes into effect and I infer p. However, subjectively what seems to happen is that I operate with a statement that says to infer p UNLESS ..., where the blank is filled in by all the IF THEN statements. This statement may be logically equivalent to the complex IF THEN ELSE statement, but it seems functionally different, in that I don't seem to check to for the fulfillment of the UNLESS clause.

An analogy with intentions might illuminate the difference. If I conditionally intend to do something, my doing it is dependent on the fulfillment of a certain condition. For example, my intention to attend this conference was conditional on getting travel funds. Compare that with a categorical intention, such as to eat dinner this evening. I certainly don't intend to eat dinner this evening no matter what. Something more exciting could come up. I could be bribed to go hungry. Yet such possibilities do not seem to be part of my intention in the way that getting travel funds was part of my intention to attend this conference.

I don't mean to suggest that there is nothing going on in me to check for contrary considerations. Maybe, as in certain AI programs, I have demons to do the job. That idea goes back to Socrates, who said,

A sort of voice [daimon] sometimes comes to me.
It never tells me what to do but only dissuades
me from doing what I am proposing to do. - Apology 31D

Still, there is the question of how, in real-life situations, the right demon goes to work and how, moreover, I get access to his findings. I don't know how these demons are organized or how particular ones take control, but we need some explanation of how cognitive processes that normally go quickly and smoothly are sometimes interrupted (see Pylyshyn 1979), and appropriately complicated, by special considerations that come to mind. The problem is not merely, as is often said, that we can be surprised or startled by something whose absence we did not explicitly expect, but also that this generally occurs just when it should. A similar phenomenon occurs in the context of skilled activity. Things are going along smoothly and then something doesn't feel right, and we pay special attention to what we're doing.

I don't have time to review the recent work in AI that is relevant here, but it is more promising than you might suppose, especially if you have been influenced by latterday phenomenologists like Dreyfus (1979) and Searle (1983). These critics of AI claim that what they call the "background" (which is what people in AI recognize as generating the frame problem) undermines the representational/computational theory of mind. Even Jerry Fodor is pessimistic about the prospects for simulating global cognitive processes (1983, 129), partly

because he suspects that nonpropositional systems of representation like Minsky's (1974) frames are not genuine but merely notational alternatives to propositional systems (1983, 116). However, some of the recent work alluded to above (e.g. Lenat 1982, 1983 and Aikins 1983) imposes enough structure on its systems, including modifiable heuristics, as to suggest that these are genuine alternatives. Besides, I know of no good reason to suppose that mental representations are propositional, at least if that means that they are couched in some form of natural language like English or in some sort of formal language like the predicate calculus. Frames, or something like them, seem much more plausible. Individual frames contain rich sets of presumptions, stereotypes, or default assignments, call them what you will. Moreover, frame systems are organized in such a way that inferences are selectively triggered and available information is selectively utilized. If anything like the frame model applies to real people, it could make sense of the idea that in every situation we find ourselves in, we have certain expectations which are noticed only in their frustration. That is, we don't test our environment to see if our expectations are met; rather, we are startled when they are not. For example, we expect the floor to be solid and we expect the lights go on when we throw the switch. These expectations would seem to be constituted by the default values of the frame that represents the current situation. When we detect something inconsistent with one of those values, we are startled and a procedure is activated to explain why. We are startled because there is a clash between some expected value and the detected value for the same parameter. The checking procedure that is thereby activated gives rise to further expectations, as constituted by the default values in the frame that represents the information about that procedure, e.g. likely results of particular tests, which may activate still further procedures as represented by further frames.

I don't think, by the way, that the frame approach really threatens the ordinary notion of belief and other concepts of so-called folk psychology. Folk psychology contains no theory of mental representation. If something like the frame model is correct, then what we call beliefs are not represented one-by-one, contrary to what the language of thought hypothesis

suggests. As a result, the frame model blurs the distinction between core beliefs and implicit beliefs, since it leaves unclear what it is for a belief to be explicitly represented. Having a certain belief may consist in a certain slot in a certain frame being filled in a certain way. However, it could just as well consist in the fact that the person's frame system works in a certain way. For example, if, when a certain expectation is not met, he runs one particular test before running another, where there is no difference in cost between the two, that he runs the first might constitute his belief that what it tests for is more probable than what the other tests for.

Be that as it may, I don't know how well I've done so far at solving the problem of what to do next, but I'm sure I can't go wrong if I solve it now by stopping.

References

- Aikens, J.S. (1983). "Prototypical knowledge for expert systems." Artificial Intelligence 20:163-210.
- Bach, K. (1978). "A representational theory of action." Philosophical Studies 34:361-379.
- Bach, K. and R. Harnish (1979). Linguistic Communication and Speech Acts. Cambridge, Mass.: MIT Press.
- Dreyfus, H. (1981). "From micro-worlds to knowledge representation." Reprinted in Haugeland (1981).
- Fodor, J. (1983). The Modularity of Mind. Cambridge, Mass.: Bradford Books/MIT Press.
- Harman, G. (1973). Thought. Princeton, NJ: Princeton University Press.
- Haugeland, J. (1981), ed. Mind Design. Montgomery, Vermont: Bradford Books.
- Lenat, D. (1982, 1983) "The nature of heuristics." Artificial Intelligence 19:189-249, 20:31-98.
- Minsky, M. (1974). "A framework for representing knowledge." Reprinted in Haugeland (1981).
- Newell A. and H. Simon (1975). "Computer science as empirical inquiry: symbols and search." Reprinted in Haugeland (1981).
- Pylyshyn, Z. (1979). "Complexity and the study of artificial and human intelligence." Reprinted in Haugeland (1981).

Searle, J. (1983). Intentionality. Cambridge: Cambridge University Press.

Simon H. (1947). Administrative Behavior. New York: The Free Press.