# THE REASONER AND THE INFERENCER DON'T TALK MUCH TO EACH OTHER

Robert P. Abelson
Yale University

In this paper, I wish to point out an important problem which has largely escaped notice in the field(s) of natural language processing. Curiously, it is implicit in the topic title of this session, "Reasoning and Inference". The fact that two terms were used in the title instead of one implies a meaningful distinction between two different types of intelligent processing.

The dictionary definitions of reasoning and inference are not much help in making a sharp distinction, but in actual use in artificial intelligence work these are two quite different traditions, each laying claim to one of the terms. Reasoning refers to the processing of abstract propositions to reach abstract conclusions. Reasoning is formal; it is based on the use of logical or mathematical rules, or special devices in a particular problem-solving domain. Inferencing, on the other hand, refers to the processing of concrete facts to reach concrete conclusions. It is informal and "common-sensical" (whatever we quite mean by that). It includes the current vein of work on "frames" (Minsky, 1974; Winograd, 1975) or "scripts" (Schank & Abelson, 1975) or "conceptual overlays" (Rieger, 1975a) in which input patterns are matched to big memory structures from which further expectations are read out.

It is conceivable that one of these two approaches is "right" and the other "wrong". It is also conceivable that the two approaches are really the same, if only we were smart enough to see how. Much more likely, mechanisms of (at least!) two rather different general types can coexist within the same intelligent system, each taking over the processing burden in its appropriate contexts.

The question I want to raise concerns <u>the interface between reasoning and inferencing</u>. I will initially explore examples of this interface in the human mind, and then comment with respect to artificial intelligence. My very tentative conclusions will be that psychological reasoning and inference processes are relatively insulated from one another; that this has unfortunate, or at any rate peculiar consequences for human thinking; that it is tempting to want to design artificial intelligence systems wherein the reasoner and the inferencer talk to each other more, although it is very unclear how best to do this; thus this problem may constitute a major agenda item for artificial intelligence in the next decade.

## <u>Mixing abstract and concrete information</u>

In the course of human affairs, there often seem to be cases where lip service is paid to abstract values (say, racial integration) but behavior is by contrast responsive to concrete concerns (say, being fearful for one's own children). As political scientist Robert Dahl puts it, "[It is] a common tendency ... of mankind ... to qualify universals in application while leaving them intact in rhetoric." On the face of it, this contrast appears hypocritical. More charitably, however, it may in some cases represent a natural difficulty in applying abstract principles to concrete cases. Recent experimental studies have demonstrated such a difficulty very strikingly.

Several studies by Kahneman and Tversky (1973) have dealt with the use of abstract "base rate" information. Subjects are given a personal background description of an individual and are asked to rate the likelihood that he chose one or another of two careers. For example:

> Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.

This background biography is written so as to suggest strongly the stereotypic inference that one occupation is much more likely than the other (above, engineer).

Now suppose that base rate information is introduced. Some of the subjects are told that the given biography was randomly sampled from a population of individuals with a fixed ratio of the two career choices, say 70% lawyers and 30% engineers. Rationally, this base rate information ought to make the more densely represented occupation more likely to be the correct characterization of any single individual. Yet it turns out that subjects given base rate information make occupational predictions no different from subjects given no base rate information or contrary base rate information. That subjects might not understand the meaning of the base rate information cannot be the explanation of this irrational result. A separate group of subjects given only base rate information and no biography correctly applies the statistical odds from the population proportions to the single case.

The explanation for this effect seems to be that the concrete impression from the biography subjectively wipes out the abstract statistical information. It is as though the inferential program scanning the biography as a single case has no way to borrow from or communicate with the reasoning program which scanned the base rate information. This explanation gains credence from the further finding that the biography doesn't necessarily have to be strongly stereotypic to wipe out the influence of base-rate information.

At a recent conference at Carnegie-Mellon on "Cognition and Social Behavior", papers by Abelson (1975a), Slovic, Fischhoff, and Lichtenstein (1975) and Nisbett, Borgida, Crandall, and Reed (1975) discussed several related phenomena. Nisbett et al. gave a number of striking experimental findings, and the following powerful anecdotal example:

"Suppose you wish to buy a new car and have decided that ... you want to purchase ... either a Volvo or a Saab. As a prudent and sensible buyer, you go to Consumer Reports, which informs you that the consensus of their experts is that the Volvo is mechanically superior, and the consensus of the readership is that the Volvo has the better repair record. Armed with this information, you decide to go and strike a bargain with the Volvo dealer before the week is out. In the interim, however, you go to a cocktail party where you announce this intention to an acquaintance. He reacts with disbelief and alarm: 'A Volvo! You've got to be kidding. My brother-in-law had a Volvo. First, that fancy fuel injection computer thing went out. 250 bucks. Next he started having trouble with the rear end. Had to replace it. Then the transmission and the clutch. Finally he sold it in three years for junk.' The logical status of this information is that of the N of several hundred Volvo-owning Consumer Reports readers has been increased by one, and the mean frequency-of-repair record shifted up by an iota on three or four dimensions. But anyone who maintains that he would reduce the encounter to such a net informational effect is either disingenuous or lacking in the most elemental self-knowledge."

If we accept this example as indicative of a general phenomenon, it may seem to imply that people do not use statistical information properly. However, that's not quite it, because if the protagonist of the story had not gone to the cocktail party, he would have used the Consumer Reports information (properly) to buy a Volvo. Rather, like the experiment with the biographies, it appears that there is an inability to translate information from one mode to the other. If you have an abstraction (a good statistical repair record) without direct episodes to instantiate it, and then in another context you are given an episode implying but not by itself proving the contrary abstraction, it is not so easy to put these pieces of information together. You cannot comfortably add the new episode to the instance set stored under the old abstraction; it's the only instance, and it doesn't support the abstraction. There is no commensurability of information, either at a concrete or at an abstract level. How would a cleverly designed artificial knowledge system cope with such a case? I do not know.

Consider another psychological phenomenon involving more reasoning than the above example, but still pitting abstract against concrete information. In the lingo of social psychology's "attribution theory" (Jones, et al., 1972), both "consensus" information and "distinctiveness" information are useful in assigning causal responsibility for an event. Given a statement about an actor's response to a stimulus ("Mary ran away from the dog"), it is possible to locate the cause either primarily in the actor ("Mary was fearful") or in the stimulus ("The dog was frightening"). From simple reasoning (setting aside complicating factors) it follows that if the given actor responds non-distinctively to other similar stimuli ("Mary is afraid of other dogs") and there is non-consensus from other actors toward this stimulus ("No one else is afraid of this dog") then it is something about Mary which has caused the event. On the other hand, if the given actor responds distinctively to other stimuli ("Mary is not afraid of other dogs"), and there is consensus from other actors toward this stimulus ("Everyone else is afraid of this dog"), then it is something about the dog which has caused the event.

In a formal logical sense, consensus and distinctiveness information are symmetric and equivalent. However, it has been found in questionnaires posing events and asking for attributions (McArthur, 1972) that distinctiveness information (what this actor does with other stimuli) is more powerful than consensus information (what other actors do with this stimulus). In fact, consensus information is curiously weak in a number of contexts.

Miller, Gillen, Schenker, and Radlove (1973) asked college students to read the procedure section of the classic Milgram (1963) study of obedience, in which subjects were found willing to administer overwhelming doses of electric shock to another person. Half the subjects were given the actual data of the Milgram study showing that all subjects administered substantial shocks and a clear majority continued throughout the experiment, even beyond the point of apparent danger to the victim's life. The other half of the subjects were left with the naive--and typical--expectation that such behavior would be quite rare. Then all subjects were asked to rate two individuals, both of whom had administered maximal shock, on eleven evaluatively loaded trait dimensions such as warmth, aggressiveness, etc.

Social psychologists have learned to interpret the apparently vicious behavior of Milgram's subjects as being due to the subtly very powerful situational pressures toward obedience, pressures with universal effect transcending personal values. If everybody does something bad, it is not as logical to fault the given individual who does it as it would be if he were one of only a few individuals who did it. Nevertheless, logic does not prevail in this judgment. The Miller et al. subjects given the true consensus information judge the two shock-givers in virtually equivalently wicked terms as do subjects not given this information. Somehow subjects are not able to adjust their concretely moralistic interpretations to take account of the base rate of the behavior in question. The value inference is impervious to statistical reasoning.

It may seem that these examples are special in some way, and that perhaps the reasoner and the inferencer do converse in other contexts. I want to turn now to another, rather different context which the artificial intelligence worker will find somewhat more familiar, but where the same puzzle arises.

How abstractly should a planner plan?

One of the vexing issues in artificial intelligence which everyone is aware of but not eager to discuss, is the lack of generality of problem contents. There is a kind of tacit agreement that it is OK for everyone to define his own arena or table top or conceptual corpus, and do his thing within that context without worrying yet about extrapolation.

I can understand the arguments for this permissive attitude--the necessity to start somewhere, the danger of doing nothing by worrying about everything, and so on--yet somehow it still bothers me. Thus I note that Terry Winograd's (1972) tour de force with the "blocks world" is applauded for its cleverness, but no one to my knowledge has tried to generalize any of the mechanisms to other problems.

Yet other problems arise which seem somehow very closely related. For example, Charniak's (1975) paper at this conference raises questions about the "supermarket frame". The shopper must get items to the checkout counter with the sometime instrumentality of a shopping cart. It seems clear that there is an analogy between the planned transport of a grocery item to a checklist counter, enabled by the prior placement of that item in a mobile cart, and the planned transfer of a block to an empty spot on a table top, enabled by the prior grasping of that block by a mobile robot arm. To be sure, there are some differences in side constraints (Shrdlu's arm can only hold one thing at a time, while the cart can hold many), but it is the similarities which interest us, even at this very simple analogic level.

Is there any way to take advantage of such similarities in the design of a more abstract problem understander? Yes, in my view. In a recent paper (Abelson, 1975b), I attempted to develop a set of intention primitives as building blocks for plans. Each primitive is an act package causing a state change. Each state change helps enable some later action, such that there is a chain or lattice of steps from the initial states to the goal state the actor desires.

These act primitives we call "deltacts". They are distinguished by the state they change, and are notated by the symbol Δ preceding a state name. There are nine deltacts in my system, designed (hopefully) such that they could be applied to almost any natural world content. (Some of these deltacts make cameo appearances in a Schank (1975) paper in this volume, embellished with his idea of planboxes, and with a new deltact added. He and I have been trying to work out ways to graft the deltacts into an expanded Conceptual Dependency formalism). For purposes of the present paper, it is not necessary to explain the entire set of deltacts, but merely to focus on a portion of the machinery needed for the robot-arm and shopping basket cases.

Consider first the abstract goal (or subgoal) that an object X be located at a certain place Y. In our abstract system, the achievement of this goal requires the deltact ΔPROX, a change in the proximity relations of an object. This deltact has five arguments: the actor A, the object X, the object's starting place Z, the final place Y, and the means M. In turn, this deltact has a set of enabling states, which may have to be achieved by other deltacts. The Abelson (1975b) paper considers the general case where the ΔPROX may be achieved via a "carrier system" (say, an airplane) run by agents other than the main actor. If we ignore this unnecessary complication (along with one other) here, the enabling states for ΔPROX may be listed as follows, where "I" denotes an instrumental device (such as a shopping cart) used for means M.

a) PROX(A,Z)--the actor must be at the starting point;

b) PROX(X,Z)--the object is at the same starting point;

c) HAVE(I,A)--the actor must have the instrumental device;

d) UNIT(X,I)--the object must be in a "unit relation" with the instrumental device;

e) OKFOR(I,M)--the device must be in good condition for the transportation means.

These enablements are very general. They must be satisfied no matter what the content. Perhaps the level of generality is too high, so that some of the states such as OKFOR may have a feeling of kluge about them (as Rieger (1975b) wonders), but this is not a problem for other states such as UNIT which have rather clean properties. (There are two variants of UNIT: nested ("in", or "on"), and joined ("with"). Various nice axioms characterize objects in UNIT relation, for example, common fate from ΔPROX; transitivity of nesting, etc.)

Concretizing these conditions to the two applications:

Robot arm moving block: The arm must be at the place where the block is; the arm must have its "hand" (satisfied by definition); the object must be put in unit relation with the hand; the hand must not be damaged (satisfied by definition.

Shopper moving an item: The shopper must be at the place where the item is; the shopper must have (say) a cart; the item must be put in unit relation with the cart; the cart must be all right for whatever it is that carts do (roll).

In placing these two applications in parallel under a common abstract rubric, we clearly see their similarities and differences. The robot problem is simplified because its hand is always there, undamaged. To satisfy the remaining two states enabling the desired ΔPROX, it is necessary to get the arm to where the block is (another ΔPROX, or in Winograd's terms, MOVE), and to join the arm to the block (a ΔUNIT in our terms, or GRASP in the original). The supermarket problem is one more complex because the shopper may not have a cart, or may have a damaged one. If he doesn't have one, he must execute a ΔHAVE, and if it is damaged, he might execute a ΔOKFOR (fix it) or a ΔHAVE of another. Additionally, he must get to where the item is (ΔPROX) and nest the item in the cart (ΔUNIT).

Each of these prior deltacts has its own abstract list of enablements. For a ΔUNIT, these include a couple of OKFORs which for grasping are not so klugey: for the robot, they map into the conditions that the arm be empty and the block have a clear top. For a ΔHAVE (which can be achieved either by taking the object or having someone give it to you), the enablements include other PROX and UNIT conditions, etc.

With a well-designed system of deltacts (which mine probably is not, because of various loopholes), it ought to be possible to have a planner (or an understander of the plans of others) reason appropriately at a level well above the specific set of content bindings of a particular problem frame or script. Apart from the academic theoretical interest of such an abstract system, would it ever be useful to human or artificial intelligence working in a given practical context?

Well, suppose something went wrong in the execution of the usual plans in the context. Suppose, for example, that Shrdlu's hand were damaged so that it couldn't grasp the smallest blocks--could it ever move them? Or suppose that the shopper absolutely couldn't find a cart.

In these cases, one could imagine the inferencer (say, of the shopper) contacting the reasoner for help, "Do something! Get through your such-thats and use a theorem or something. I'm in trouble here without a cart." And the reasoner might say, "Well let's see, you ask for an instrumental device such that you can have it and it's OK for forming unit connections with a number of these--what did you call them?--grocery items, in order to move them a short distance? Well, how big and heavy are these grocery items, and how soon do you want an answer?" And three minutes later the reasoner would come back and say, "Since it's an emergency, I'll be more specific than I usually like to be. I recommend something flat, with a big surface area, that you already have and that could be pushed or pulled by hand. Have you got anything on your person?"

"Well, I have this overcoat."

"Oh, good. Yes, use your overcoat to pull the grocery items along the floor to the check-out counter."

"What?" (Horrified) "But the overcoat will get filthy!"

"Sorry. That's not my department. That's a low-level inference. I thought you wanted high-level reasoning."

An alternative fantasy is that the reasoner might be so busy working out abstract puzzles with its abstract mechanisms (say, running through a UNIT exercise on the Towers of Hanoi problem) that it wouldn't be interested in troubling with a silly applied problem. "Go away and don't bother me! I'm thinking about the Towers of Hanoi. What do I know from groceries?!"

## Reprise

The argument I have stated as a devil's advocate goes back at least to William James, writing in 1890 of the several "worlds" of the mind:

> "Every object we think of gets at last referred to one world or another...It settles into our belief as a common-sense object, a scientific object, an abstract object, a mythological object, an object of someone's mistaken conception, or a madman's object, and it reaches this state...often only after being hustled and bandied about amongst other objects until it finds some which will tolerate its presence and stand in relations to it which nothing contradicts. The molecules and ether-waves of the scientific world, for example, simply kick the object's warmth and color out, they refuse to have any relations with them...[The] world of classic myth takes up the winged horse;...the world of abstract truth, the proposition that justice is kingly, though no actual king be just. The various worlds themselves, however, appear (as aforesaid) to most men's minds in no very definitely conceived relation to each other, and our attention, when it turns to one, is apt to drop the others for the time being out of its account. Propositions concerning the different worlds are made from 'different points of view'; and in this more or less chaotic state the consciousness of most thinkers remains to the end. Each world whilst it is attended to is real after its own fashion; only the

reality lapses with the attention." (James, 1950, p. 293).

Is James right? If not, wherein? If so, is that how artificial intelligence--which possibly has design options not available to the human mind--would like to keep it?

## References


Abelson, R.P. Script processes in attitude formation and decision-making. Presented at the Eleventh Carnegie Symposium on Cognition, Carnegie-Mellon University, April, 1975. (a)

Abelson, R.P. Concepts for representing mundane reality in plans. In D. Bobrow and A. Collins (Eds.), Representation and Understanding. New York: Academic Press, 1975. (b)

Charniak, E. Organization and inference in a frame-like system of common knowledge. In Theoretical Issues in Natural Language Processing. Proceedings of conference at the Massachusetts Institute of Technology, June, 1975.

James, W. (1890) Principles of Psychology. Vol. 2. New York: Dover edition, 1950.

Jones, E.E., Kanouse, D.E., Kelley, H.H., Nisbett, R.E., Valins, S., & Weiner, B. Attribution: Perceiving the Causes of Behavior. General Learning Press, 1972.

Kahneman, D., & Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80, 237-251.

McArthur, L.A. The how and what of why: Some determinants and consequences of causal attribution. Journal of Personality and Social Psychology, 1972, 22, 171-193.

Milgram, S. Behavioral study of obedience. Journal of Abnormal and Social Psychology, 1963, 67, 371-378.

Miller, A.G., Gillen, B., Schenker, C., & Radlove, S. Perception of obedience to authority. Proceedings, 81st Annual Convention of the American Psychological Association, 1973, 127-128.

Minsky, M. A framework for representing knowledge. MIT AI Memo 306. June, 1974.

Nisbett R.E., Borgida, E., Crandall, R., & Reed, H. Popular induction: Information is not necessarily informative. Presented at the Eleventh Carnegie Symposium on Cognition, Carnegie-Mellon University, April, 1975.

Rieger, C. Conceptual overlays: A mechanism for the interpretation of sentence meanings in context. To appear in the Proceedings of the 4th International Joint Conference on Artificial Intelligence. Tbilisi, 1975. (a).

Rieger, C. The commonsense algorithm as a basis for computer models of human memory, inference, belief, and contextual language comprehension. In Theoretical Issues in Natural Language Processing. Proceedings of conference at the Massachusetts Institute of Technology, June, 1975. (b)

Schank, R.C. Using knowledge to understand. In Theoretical Issues in Natural Language Processing. Proceedings of conference at the Massachusetts Institute of Technology, June, 1975.

Schank, R.C., & Abelson, R.P. Scripts, plans, and knowledge. Prepared for presentation at the 4th International Joint Conference on Artificial Intelligence. Tbilisi, 1975.

Slovic, P., Fischhoff, B., & Lichtenstein, S. Cognitive processes and societal risk taking. Presented at the Eleventh Carnegie Symposium on Cognition, Carnegie-Mellon University, April, 1975.

Winograd, T. Understanding Natural Language. New York: Academic Press, 1972.

Winograd, T. Frame representations and the declarative/procedural controversy. In D. Bobrow & A. Collins (Eds.), Representation and Understanding. New York: Academic Press, 1975.