

Notes on a Cognitive Process Model of Bargaining (Type IV A)

Paul Diesing

Working Paper no. 7.

One moves toward a cognitive process model of bargaining by dropping the assumption of rationality that is made in utility and game models of bargaining. Consequently, it is necessary first to examine this assumption to see what part it plays in these models, what modifications or adjustments of the assumption are possible within the models, what considerations can finally lead to its abandonment, and what requirements must be met by a replacement.

In utility models the assumption is that 1) the bargainers have fully worked out preference schedules, in the sense that when they are presented with any two alternatives they know which one they prefer; 2) these preference schedules do not change during bargaining; 3) each bargainer knows or can find out the preference schedule of the other. Game models assume in addition; 4) there is a fixed and finite set of alternative strategies available to each bargainer; 5) the consequence of each strategy for both players is known. These assumptions vary somewhat from one theory to another, but the above set is representative.

Of course all these assumptions are unrealistic to some extent, and that is precisely why they are valuable. If our bargaining theory were completely realistic it would be as complex as the real bargaining processes it describes, and indeed a great deal more complex since it would clarify and make explicit all the hidden interconnections that occur in reality. The point of theory is to simplify reality in such a way that it becomes manageable. These assumptions are justified if they do that.

The function of the above five assumptions is to focus attention on certain essential characteristics of bargaining that are thought to constitute the essential bargaining problem. These essential characteristics are 1) conflict of interests and 2) interdependence of outcomes. The bargaining problem is how

to achieve a preferred outcome when all outcomes depend on the acts of an opponent with conflicting preferences. The fact that many other things occur in bargaining should not distract one's attention from this central problem. To be sure, bargainers are sometimes confused or uncertain about their preferences, but that is not a bargaining problem. To be sure, they change their preferences and develop new interests, but that is not a bargaining process (so the argument for these assumptions goes). To be sure, they are often doubtful or mistaken about opponents' preference schedules and strategies and about the consequences of their own actions, but that is a universal human problem, not a bargaining problem. To be sure, alternatives may not all be known or given at the start of bargaining, but they do not become part of the bargaining problem until they become known. The assumption of rationality abstracts from all these human problems to focus on the essential bargaining problem. It states that if bargainers are not distracted by ignorance, confusion, uncertainty, or creativity, but immediately choose the best strategy open to them; then what happens when A's best is bad for B and vice versa?

The assumption of rationality is, I believe, justifiable in three kinds of circumstances. 1. It is justifiable if the factors assumed away are of minor importance and do not affect the outcome significantly. If preferences are reasonably clear and do not change much, if alternatives are pretty well given, and if uncertainty and ignorance can either be pretty well dispelled during bargaining discussions or can cancel each other out or do not really change the outcome, then it makes sense to treat the bargainers as rational. If you think your case is of this sort, a Type IV model represents unnecessary complications for you. 2. It is justifiable as an idealization or ideal type. Such a model describes the outcome that would occur if both bargainers were rational; the actual outcome will deviate to just the extent that one or the other falls short of rational behavior. For instance, the effect of specific mis-estimations of

outcomes or of opponent's preferences or choice probabilities can be estimated pretty exactly. The effect of threats by one person can be estimated (Kent). This tactic works if the deviations from rationality produce simple, direct deviations from the predicted outcome. However, if both bargainers deviate considerably from rational behavior the deviations may interact and produce effects that cannot at all be understood by reference to the model. This occurs, for instance, if each bargainer is trying to deceive the other and both are aware of this but each thinks he can outwit the other; the outcome of this contest cannot be calculated by measuring deviations from the ideal of omniscience.

However, one can still use the rationality assumptions when there are extensive deviations, if one is willing to give up exact description and prediction. Bargaining then becomes an "art" which cannot be scientifically described or understood because it deals with imponderables--that is, factors which have been excluded from one's model by the rationality assumptions. For instance, Ikl'e holds that changes of utilities are an essential part of the bargaining process; but since all changes of utility are deviations from rationality they cannot be modelled but belong to the art of bargaining. They can be described only negatively and vaguely as deviations from a utility model. Similarly, the discovery of a new strategy can be described only negatively by means of a payoff matrix; it becomes an unpredictable, creative act which goes beyond the matrix and changes it.

3. The assumption of rationality is justifiable if there are other models of the excluded processes and if these models can be hooked on to a bargaining model without changing its internal workings. This works if the other models are limited to determining parametric values, for instance payoff values, but do not affect the bargaining process itself.

The assumptions of rationality must be discarded if the excluded processes are important, if their effects are too complex to be describable simply as a deviation from the rational outcome, and if they operate within the bargaining

process itself.

A replacement must provide a systematic account of the processes excluded from utility and game models by the rationality assumption. This includes; 1. information processing and 2. decision making. Under 1. is included (a) interpreting incoming information; (b) the development and modification of a bargainer's image or expectations about his opponent; (c) remembering the results of a and b. These processes serve as a substitute for the "omniscience" part of the rationality assumption. Under 2. is included (a) marshalling relevant value-criteria or constraints, which may take a variety of forms; (b) devising possible strategies one by one; (c) testing each successive strategy by the criteria until a satisfactory one is found. These processes serve as a substitute for the fictitious fixed preference schedules and fixed alternative strategies in the rationality assumption.

A semantic difficulty may hinder some from approaching the above task. It may be supposed that if one rejects the assumption of rationality one is then left only with irrationality: confused and contradictory preferences, misperception and misunderstanding, spontaneous creativity. These are, however, negative terms which indicate that one is thinking in terms of deviations from a rationality model. When one models these processes positively one uses positive terms; one speaks of diverging interpretations rather than misperceptions heuristic search rather than spontaneity, etc.

SOURCES

Coddington (1968), building on Cross (1965) has proposed a bargaining model centered on information processing and decision making. His proposed model includes 1) expectations or images about the opponent, 2) decision on a bidding move, based on those expectations, 3) interpretation of results and adjustment of expectations. Then the cycle repeats until opposing bids match. Unfortunately Coddington does little more than suggest the model and argue that it would be a

good one. Jervis provides material for a model but no model. Consequently, I have had to go elsewhere for systematic ideas. For information processing I have gone to cognitive psychology, including such works as Norman, Recognition and Memory, (1969), Reitman, Cognition and Thought (1965), Sayre, Recognition (1965) Feigenbaum and Feldman, Computers and Thought (1962). For decision-making I have yet to see an improvement on Newell and Simon's GPS model, as modified by Cyert and March (1963) and with extensions suggested by Reitman (1965).

These sources mean that the model will be a computer model. This is a disadvantage because construction of a computer model is a terrible nuisance, especially for a novice like me. However, we are unlikely to get to the stage of actually writing a computer program. Instead, our concern must first be with the general outlines of such a model, and these can be worked out and modified at the verbal level. Other and wiser theorists can take over from there.

INFORMATION PROCESSING

This includes A) interpreting received messages, both signals and observations of an index; B) remembering interpreted messages, including storing and retrieving from memory; C) modifying images or expectations on the basis of interpreted messages.

A. Interpretation. In working paper No. 5, p. 16, I described interpretation as a matching process, in which a message is compared with successive items of an expectation list until a satisfactory match is found. This conception of interpretation, or recognition, was current in computer models of the late 50's and early 60's, but I find that things have now changed. Interpretation, or recognition, is now seen as an active process of producing a message that matches the message to be interpreted and then assuming that the source has produced the message in the same way. This is called "analysis by synthesis". For instance, suppose Germany receives a message from England. Germany retrieves its image of England from memory and asks, "By what process might England have

produced that signal?" It then activates the England-image by imagining the British to be composing a message on the topic discussed in the actual incoming message. It produces, one after another, messages that England could have sent on that subject. Each message is tested against the received message until one of them matches. It is then assumed that the intention or strategy that produced the successful matching message is the actual intention underlying the original message.

The same process of interpretation works equally well on signals, signs and indices, though there may be differences of detail.

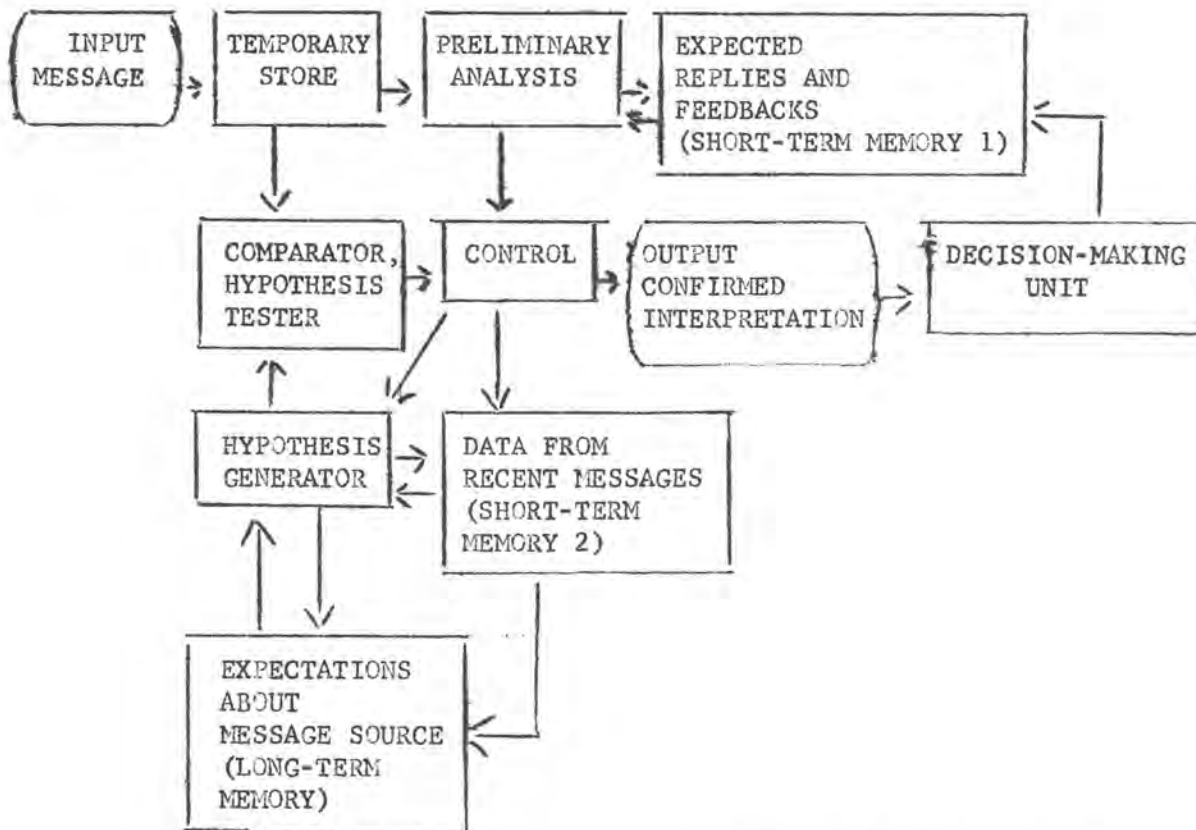
For a more specific example, consider Sazonou's messages of August 26 and 27 to Germany: a) Russia is not mobilizing yet; b) if Russia mobilized, it would be against Austria only, not against Germany; c) Russia desires discussions with Austria leading to a compromise. The German image of Russia is that it is weak, militarily unprepared, on the verge of internal revolution, without dependable allies, and likely to back down in a crisis. The latter two elements are based on Russian and French actions in 1908-1909. How would such a country produce the above signals? The implicit suggestion of mobilization in a and b cannot be produced by a militarily unprepared country, so the Germans are unable to produce a matching signal from their image. This means that they do not recognize the implicit threat, they do not notice it. They can produce the message as a bluff but not as a serious threat. But c can be produced out of the "likely to back down" elements. A country that wishes to back down could express this wish in a signal that it desires to compromise. Consequently c is interpreted as expressing an intent to back down, and incidentally as confirmation of the "likely to back down" image.

The two theories of interpretation, the matching theory and the active "analysis by synthesis" theory are not as different as they seem at first. The main difference is that in the active theory there are two matching operations, the first a general preliminary match and the second a hypothesis-testing match.

Consequently it is possible to incorporate the older theory in the newer one by allowing some bypassing of the second matching operation. This is particularly appropriate, I think, when the interpreting country is definitely expecting or searching for a certain kind of message. It may be expecting a reply, yes or no, to its own message; or if it is carrying out an active strategy it may be interested only in feedback about whether or not that strategy is succeeding. In such cases it may filter out everything except the desired information; in extreme cases it may filter out everything except success reports, for instance if the information receptors dare not displease their master back home. I have incorporated these ideas in the detailed routines below.

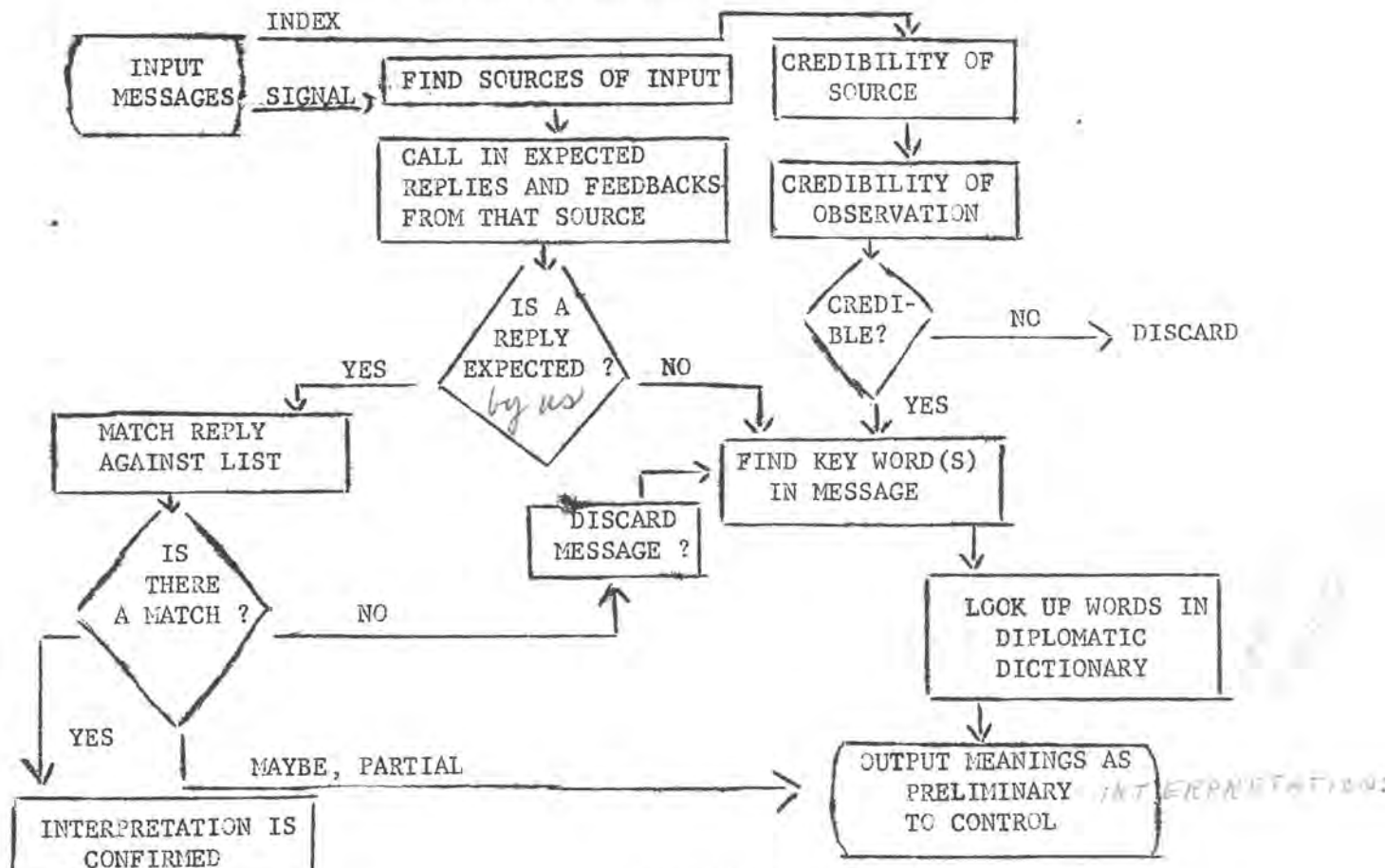
In general, the whole theory is basically a working out of the old common-sense idea that we interpret an ambiguous message in terms of what we expect the sender to say or do; and that when we are actively working on the environment we are interested only in information about how well we are doing.

PROGRAM CONTROL CHART

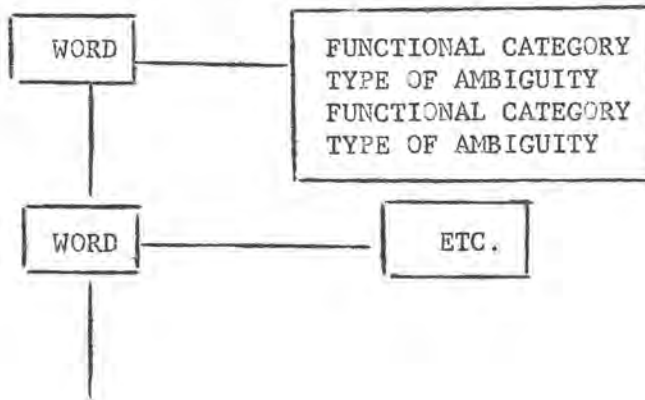


Adapted from M. Halle and K. Stevens,
 "Speech Recognition". IRE Transactions
on Information Theory, 1962, pp. 155-159.

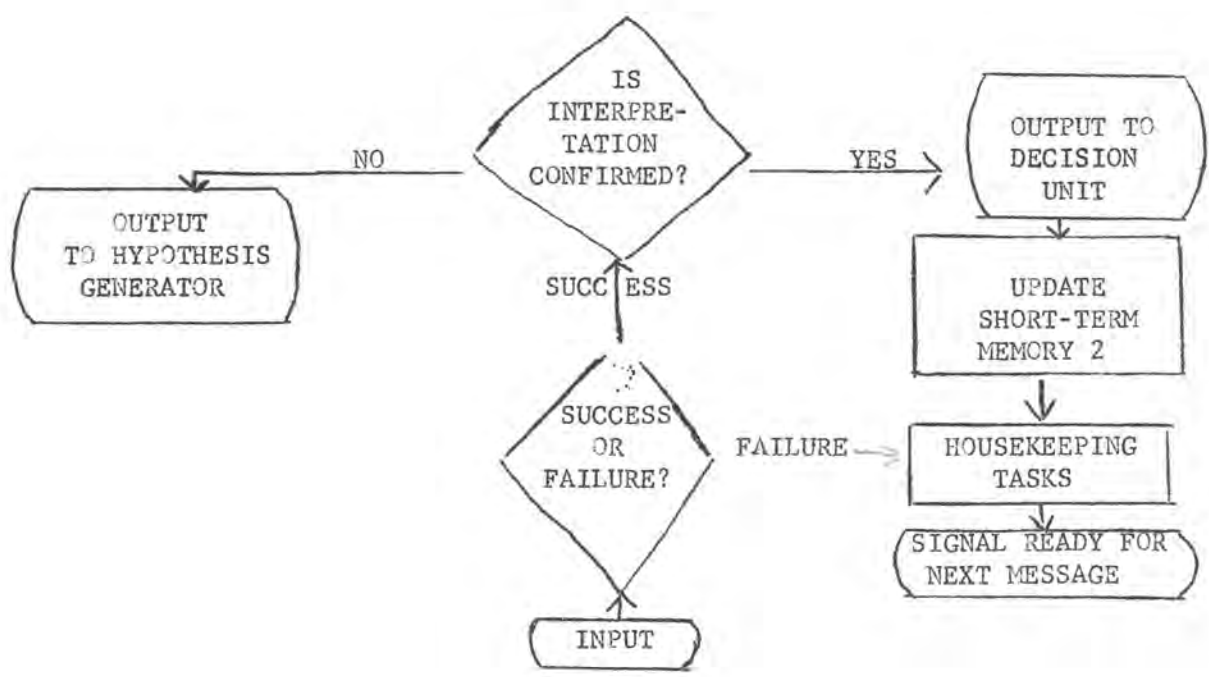
PRELIMINARY ANALYSIS FLOW CHART



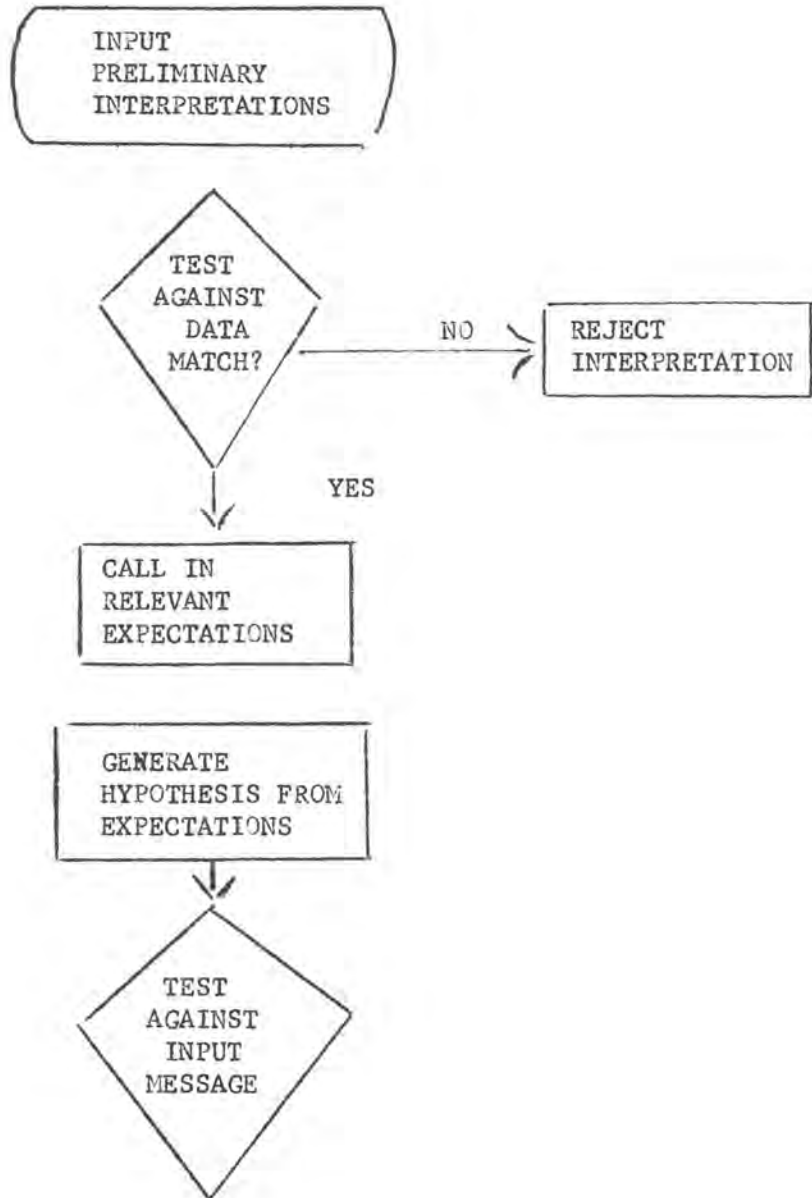
DIPLOMATIC DICTIONARY



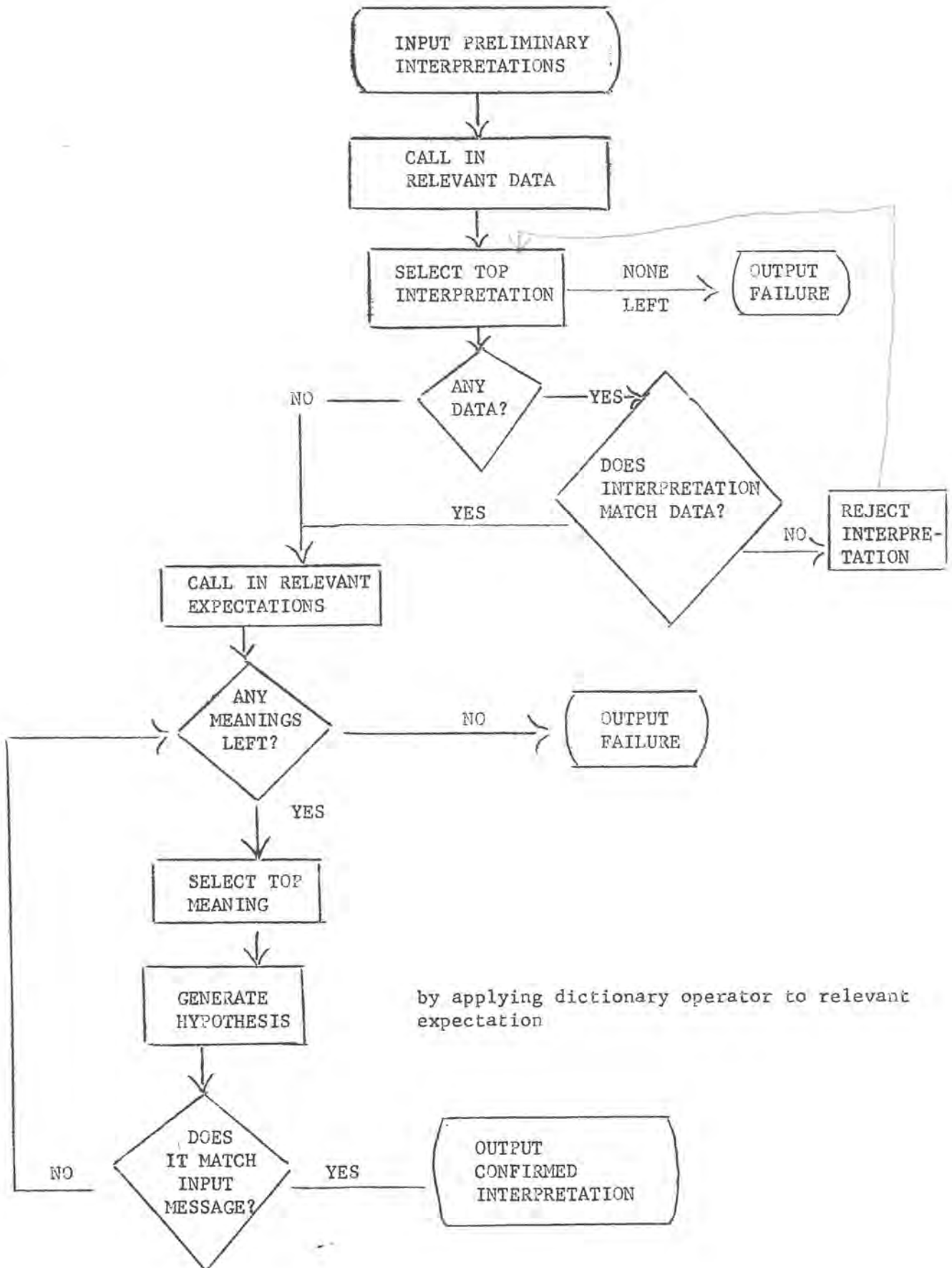
CONTROL



HYPOTHESIS GENERATOR



HYPOTHESIS GENERATOR (DETAIL)



B. Long-term memory; C. Modifying expectations.

I suggest that the process of modifying expectations in the light of new data is essentially a storage problem in which data are transferred from short-term to long-term memory. The reason is that expectations are stored in l. t. ^{long term} memory and data are combined with them in the storage process. Consequently B and C can be treated as a single topic, namely memory storage and retrieval.

Norman (1969) states three generally (?) accepted principles governing memory storage and retrieval.

1. "In order for material to be stored in secondary memory, it must be integrated within the existing organization." (p. 135)
2. The integration of memory increases over time. (p. 137)
3. An item is retrieved from memory by activating a whole organized unit and running through it until the desired item appears.

Let us see what implications these three principles have for the system of expectations. First, data about recent messages are stored in short-term or primary memory. This means that they are soon forgotten unless they are transferred to long-term memory. But the transferral is successful only if the new data can be integrated with old expectations. This may involve some reorganization of expectations; but if new data directly contradict old expectations they will be forgotten. For example, if a supporter of the Vietnam war hears of American atrocities--supposing that he could even interpret a message that way, which is unlikely--he will soon forget them, while an opponent will remember them. If a foreign minister, who sees himself as honest and trustworthy, sends out a few deceptive messages--supposing that he could even interpret them as deceptions, which is unlikely--he will soon forget that he ever deceived.

Instead of forgetting contradictory data, a person may modify them to fit the existing image and will remember them in that form. The deception will not be remembered as a deception, but as something else. This phenomenon is the

basis of the familiar experience of hearing opposing sides give contradictory reports about some past period of conflict. Nobody is lying; they are just remembering.

If the existing image is loosely integrated a new datum may activate powerful emotional forces that drastically reorganize the whole image, turning it upside down or reversing its basic charge from plus to minus or minus to plus. This, however, is rare.

Second, the integration of images increases over time. I suspect that this is primarily a process of cognitive balancing, in which the plus charges and the minus charges are all collected, segregated, and balanced, (See work by Abelson, Brehm, Festinger, etc.). What this means is that the image of an ally tends to get ever more positive and that of an adversary ever more negative. However, there is almost always a split within the image of the ally, with the dominant forces positive and the minority forces negative; the minority force is held responsible for all negatively-charged acts of the ally. There is a similar, opposite split within the image of the adversary. When the split image is reversed or turned upside down the adversary becomes an ally in disguise and vice versa.

Third, when a re-organization or even reversal of an image occurs, some parts of the image are likely to be dissociated from the new organization. For instance, if the new image is positive, there may be some left over negative items that could not be reconnected to the main negative component. These leftover items remain in memory but can no longer be retrieved or "remembered." Retrieval consists of activating the relevant part of an image and running through it until the desired item appears; but this means that dissociated items cannot be retrieved. This sort of forgetting was first experimentally demonstrated and explained by Feigenbaum and Simon with their EPAM model (1961).

The dynamics of this sort of model cannot be determined without using a

computer, but some simple guesses are possible. It seems obvious that outcomes are determined primarily by the rules for long-term memory, secondarily by the thoroughness of the preliminary analysis, and perhaps slightly by other elements such as the diplomatic dictionary. In broad empirical terms this means that the effect of diplomatic communication in a crisis depends primarily on the differences among the countries' images of each other and of themselves as they have developed during a decade or more, and secondarily on diplomats' first quick intuitive hunches as they read a message or hear an index observation report.