[Reprinted from The American Behavioral Scientist, May 1965, Vol. VIII, No. 9]

Computer Simulation of Individual Belief Systems*

By Robert P. Abelson and J. Douglas Carroll

In this paper we sketch an attempt at computer simulation of individual belief systems. Although rooted in a series of earlier formulations (Abelson and Rosenberg [1958],[1] Abelson [1959],[2] Rosenberg and Abelson [1960],[3] Abelson [1963][4]), this complex project has not heretofore been totally outlined in print. We shall begin with several clarifying comments on the nature of our goals and the problems we have faced, thence proceeding further and further into the details of operation of our simulation.

By an *individual belief system* we refer to an interrelated set of affect-laden cognitions concerning some aspects of the psychological world of a single individual. In simulating such a system, our intention is not only to represent its structure of interrelationships, but also some of the processes by which the system maintains itself against the intrusion of new and potentially upsetting information. Our use of the technique of *computer simulation* is intended to maximize the explicitness with which we state our assumptions and the vividness with which the consequences of these assumptions are made apparent. The operation of simulated belief systems can be played out on the computer and the details scrutinized in order to refine our level of approximation to real systems.

The psychology of belief systems lies athwart the ancient philosophical battleground of whether man is basically "rational" or "irrational." Without claiming to have resolved this honorable controversy, we adopt the useful compromise position current among social psychologists that man is "subjectively rational," i.e., rational within the constraints of his own experience and motivation. This position owes much to Heider's[5, 6] analysis of "naive psychology," and can be found implicitly or explicitly in the work of many "cognitive consistency" theorists (Festinger[7]; Osgood and Tannenbaum[8]; Rosenberg and Abelson[3]). To be sure, there is still much room within this compromise position for differences in emphasis as between motivational and cognitive components of the total system. The work of Rokeach[9] and Smith, Bruner, and White[10] among others reminds us that individual belief systems are heavily determined by personality needs. Our simulation is nevertheless primarily a cognitive model, not *per se* a personality or motivational model, nor (as with many aspects of Festinger's[7] dissonance theory) an action model. The cognitive processes of our system are driven by motivational forces, but are also quite responsive to "reality" (as perceived by the individual). The achievement of a balance of forces between self-interest and reality is one of the most difficult theoretical problems in the psychology of belief systems. It is far too glib to assume that "people believe what they want to believe," since unpleasant realities must often be accepted in some form, and pleasant absurdities must often be rejected. The way in which our simulation attempts to tread this thin line is outlined later.

A second problem area pertains to the degree of mutability inherent in belief systems or, more properly, the range of tolerance for change under appropriate circumstances. Among our assumptions relevant to this question, the following three are crucial: 1) The greater the affect associated with the elements of a belief, the greater the resistance to any change in that belief or any reduction in the strength of that affect; 2) The more "closed" (i.e., massively self-consistent) a belief system, the greater the likelihood that resistive mechanisms will succeed; 3) Substantial changes can take place only when resistive mechanisms are not invoked, or when they fail.

The last assumption, which might be rephrased, *belief change and affect change are the default of resistance*, is not an empty proposition. The conditions for the use and success or failure of resistive mechanisms are precisely specified in our simulation.

Another thorny dilemma in the design of a simulation such as ours is the choice of a linguistic structure in which beliefs are to be stored for later processing. To give the system a full command of natural language is far beyond the realm of present feasibility, but a limited command of a kind of pidgin English is entirely possible. Our system has a rudimentary grammar which enables it to cope with a rather wide variety of verbal material. However, we must pay a price in both system flexibility and psycholinguistic sophistication to buy simplification.

In an interesting discussion of psycholinguistic problems, Miller[11] discusses three possible levels of analysis: syntactic, semantic, and pragmatic. Syntactics pertains to

* The work herein reported was in part supported by Grant MH-04682 from the National Institute of Mental Health, United States Public Health Service. We wish to thank Mr. William Reinfeld for carrying out the computer programming for the simulation model.

the distinction between the grammatical and the ungrammatical; semantics, between the meaningful and the meaningless; pragmatics, between the acceptable and the unacceptable. Our system has syntactic capability in terms of certain limited sentence-processing rules. However, it does not have separate semantic and pragmatic capabilities, since the former is largely subsumed by the system under the latter. To illustrate, consider the two sentences, "Lamps eat geology," and "Barry Goldwater believes in socialism." The first is semantic nonsense, whereas the second is potentially meaningful but obviously false. Our system somewhat distorts this distinction. Given sufficient information to recognize and deal with the sentence components, it would immediately adjudge the second sentence incredible, but would spend a great deal of time in futile memory search before concluding that it could not clearly regard the first sentence either as credible or incredible. A separate semantic capability would enable the system to identify nonsense immediately, but we have bypassed this design problem to concentrate directly on pragmatics.

One further remark concludes this introductory section. In the article referred to above, Miller notes, "At the present time, linguists and logicians have a rather deep understanding of syntax. . . . The general shape a semantic theory must assume is slowly emerging. Pragmatics, however, is still the wastebasket into which all miscellaneous and confusing problems are put . . ." In view of such admonitions and in the light of common sense, it is apparent that we are on pioneering ground and regard our theory as subject to considerable modification.

We now proceed to outline the details of the system.

The Memory Structure

The standard linguistic unit in which beliefs are stored is a *sentence,* consisting of a *concept* followed by a *predicate.* A predicate typically consists of a *verb* followed by a concept. Thus a sentence is in effect a concept-verb-concept (i.e., subject-verb-object) construction referring to a state of affairs which the system "believes." Since a "state of affairs" can itself be cognized as an elemental unit, every sentence in the system may also function as a concept to which other elements may be affixed to form compound sentences. For example, consider the initial sentence, "Cuba subverts Latin America" formed from the concept "Cuba" and the predicate "subverts Latin America." This sentence can in certain contexts be regarded as the noun phrase, "Cuba's subversion of Latin America." Thus a belief concerning Kremlin control of the presumed subversion could be carried by the sentence, "Russia controls Cuba's subversion of Latin America." In our terms, "Russia" is of course a concept, and "controls Cuba's subversion of Latin America" is a predicate.

This kind of packaging of one belief within another is an important feature of our system, for it permits with simple syntax the orderly storage of important conceptions about responsibility, causation, and impact involving three or more actors and actions. The network of beliefs so stored may be said to comprise the "horizontal" structure of the belief system.

For a demonstrational system presently being assembled by the first author, it appears that between 500 and 1,000 sentences are necessary to capture the basic themes and illustrations expressing the publicly stated foreign policy beliefs of a particular individual (a well-known right-winger). It should be noted, however, that a substantial portion of this corpus of sentences is devoted to definitional information rather than to beliefs *per se.* For example, consider the sentence, "Colonial powers oppose anti-colonial policies." This statement is very close to a tautology and hardly enriches the system at all. However, if this sentence were omitted, the simulated system might not "understand" the meaning of the concept "anti-colonial policies." As far as the system is concerned, the denotative meaning of an element resides entirely in structural connections of that element with other elements. Thus, if the only reference in the system to "anti-colonial policies" were, say, "The U.N. promotes anti-colonial policies," the system would be unable to use this belief meaningfully. It might just as well find stored a nonsense term in this sentence, such as "The U.N. promotes squadgy widge." This digression establishes that some of the "beliefs" of the system should technically be regarded instead as "knowledges."

Another portion of the memory structure is a "vertical" network among the elements. The basic vertical connection between elements is the *instance* relationship. The instances of a concept are those other concepts which represent "kinds of X" or "examples of X." Instances may themselves have instances, and so on.

Many instance relationships embody simple objective "knowledges," e.g., "Brazil" as an instance of the concept "Latin American nation." There may also be a subjective quality involved, as with "India" as an instance of "left-leaning neutral nations."

Compound concepts, e.g., "The Communists seek a Communist-dominated world," may also have exemplifying instances, e.g., "The Communists take over free nations." Predicates are also subject to the instance relationship; for example, "distributes atheistic propaganda" could be an instance of "promotes atheism."

The obverse of the relation "Y is an instance of X" is "X is a *quality* of Y." Thus, "Latin American nation" is a quality of "Brazil," and similarly for the other examples above. Metaphorically we conceive of a hierarchy of levels of abstraction, in which the instance relation points "down" from the general to the specific, and the quality relation points "up" from the specific to the general. The systemic significance of this entire vertical network is that it encodes the subjective codifications and categories used by the system. It is the system's particular way of parsing from reality the "who" and the "what" entering the horizontal network of who does what, what causes what, and so on.

Several other features have been incorporated into the memory structure. The most important of these concerns the affective investment of the system in its beliefs. It

is beyond our purview to simulate the inner dynamics of human affect, but it is imperative that affective influence upon cognition be represented. We have chosen the simplest big step in this direction by providing for the storage of *evaluations* of elements. Every element, whether concept or predicate, carries a signed quantitative evaluation summarizing the resultant of all positive or negative affects attaching to the element. These evaluations exert a heavy guiding influence upon certain cognitive operations of the system, and are in turn sometimes subject to change consequent upon certain cognitive outcomes. The details are sketched later.

Cognitive Processes

The present system is provided with two major cognitive processes (sub-routines, in computer programming parlance). These two processes can operate in different circumstances upon a variety of contents. One process we label the "Credibility Test"; the other, the "Rationalization Attempt." Several system capabilities are provided by these two processes, as will unfold in the course of our description of their operation.

The Credibility Test is employed by the system in two rather different ways: to screen incoming sentences for plausibility and to check the reasonableness of self-produced sentences which the system would "like to believe." The input to the process is a sentence (i.e., a concept-predicate pairing), and the output is one of three alternative judgments: that the sentence is credible, that it is dubious, or that it is incredible.

The obvious first response of the system in testing the credibility of an input sentence is to consult its memory to see if the sentence (or its opposite) is already stored. If this simple operation fails, then the Credibility Test relies upon other information in the vertical memory structure. The details of the process are quite elaborate, and only the main ideas will be sketched here.

Denote the input sentence symbolically as C-P. For the sake of a concrete illustration, suppose that the concept C is "Liberals," and the predicate P is "support anti-colonial policies." Suppose further that the system does not have in its memory either the input sentence, "Liberals support anti-colonial policies" or the contrasting sentence, "Liberals oppose anti-colonial policies."

What happens next is that a systematic search of beliefs about the instances of C will be undertaken. For each instance C_i (e.g., "Stevenson"), the system examines all beliefs C_i-P_j (e.g., "Stevenson makes flowery speeches," "Stevenson ran for President," etc.) to see whether a P_j can be found which is an instance of the original P. For example, the critical predicate here might be, "opposes Portugal on Angola," an instance of "support anti-colonial policies."

In effect, a chain of inductive evidence supporting C-P may be found by going "down" to C_i, "across" to P_j, and back "up" to P. Thus "Stevenson opposes Portugal on Angola" lends some credence to the generalization, "Liberals support anti-colonial policies." However, the location of a single such chain is not necessarily sufficient to support the generalization. It depends upon the relative number of instances of C through which such chains exist. Another instance of "Liberals" in our example might be "Earl Warren," and the system might find no example of any kind of anti-colonial policy supported by Mr. Warren. Thus the system must have some criterion by which it balances off successful and unsuccessful searches.

We have given the system the following rule: The sentence C-P is considered credible if *at least half* of the instances of C are found to be connected in a belief sentence with at least one instance P. This "majority" rule may smack of a certain arbitrariness, but it seems at least as reasonable as any other rule we might have postulated. (In a later section we report an empirical study with important implications for the choice of a rule.)

If the system fails to establish the inductive credibility of a sentence C-P, it is given the option under some circumstances of testing the inductive credibility of the opposed sentence, C-not P ("not P" is the predicate formed from P by negating the verb). Should this test succeed, the original sentence C-P is adjudged not credible. Should this test fail (or not be invoked), the Credibility Test process proceeds to undertake a somewhat more complicated search involving the qualities rather than the instances of C and P.

Search proceeds up the vertical tree and thence back down to the level of C-P. The effort is made to establish the credibility of C-P *deductively* by first establishing the credibility of several higher-level generalities. The rule for this sub-process is that at least half of the qualities of C must be found credibly related to at least one quality of P. In this latter search, the previous inductive mechanism may again be invoked.

Perhaps the further pursuit of our previous example will clarify these circumlocutions. One of the qualities of "Liberals" might be "Left-wingers" and one quality of "support anti-colonial policies" could be (in the subjective terms of a right-wing view) "mistreat U.S. friends abroad."

The credibility of "Left-wingers mistreat U.S. friends abroad" (if not previously established by the system) could be examined in terms of the instances of its components. The system would search to see whether at least half of the instances of "Left-winger" were connected to at least one instance of "mistreat U.S. friends abroad." One instance of "Left-wingers," say, "Administration theorists," might be found connected with "coddle left-leaning neutral nations," an instance of "mistreat U.S. friends abroad." A sufficient number of further instances of "Left-wingers" might similarly be connected with instances of "mistreat U.S. friends abroad," establishing the credibility of this sentence. A similar process applied to other higher-level sentences might then eventually satisfy the rule for establishing the deductive credibility of the original sentence C-P, "Liberals support anti-colonial policies."

If the rule is not satisfied, then this entire search process may under some circumstances be repeated on the opposed sentence, C-not P. If this is done, but fails, then the Credibility Test process will terminate with a signal

that it can find insufficient evidence one way or the other about C-P. As far as the system is concerned, the sentence is neither credible nor incredible.

The above description covers the essentials of operation of the Credibility Test. There are some additional wrinkles, such as mechanisms for discovering new instances and qualities by parsing compound elements, for searching two or three levels down or up in the vertical structure, and for terminating very long searches because of "fatigue," but the details will not be given. It is noteworthy that the two major sub-processes alone are capable of generating a very large quantity of subjective evidence relevant to the credibility of an assertion.

Although our description of the Credibility Test may make it appear to be a more or less neutral process innocent of the affective investments of the system, this is not necessarily the case. One fundamental use of the test is in the mechanism we label Denial. In this mechanism, the system takes a sentence C-P which is credible but upsetting, constructs C-not P, and enters the Credibility Test with the injunction to find C-not P credible if at all possible. Success in this procedure enables the system to deny C-P by means of contrary evidence.

The other fundamental process of the system, "Attempt Rationalization," is applied only to sentences which the system is motivated to "explain away." Such sentences are characterized by evaluative inconsistency or imbalance between concept and predicate, for example, a positively valued actor (concept) connected to a negatively valued action (predicate). The function of the rationalization process in these cases is to deny the psychological responsibility of the actor for the action. This can be done in three primary ways: 1) by assigning prime responsibility for the action to another actor who controls the original actor; 2) by assuming that the original action was an unintended consequence of some other action truly intended by the actor; 3) by assuming that the original action will set other events in motion ultimately leading to a more appropriate outcome.

These three mechanisms represent sub-processes of "Attempt Rationalization," respectively labeled "Find the Prime Mover," "Accidental By-product," and "Reinterpret Final Goal." Each of the three depends upon the discovery by the system of a critical element specially related to either C or P in the original sentence C-P. This critical element must be either: a concept, B, occurring in a belief sentence of the form, "B controls C"; or a predicate, Q, occurring in a phrase, "Q can accidentally cause P"; or a concept, R, occurring in "P may lead to R." The existence of such elements in the system is dependent upon some special memory storage provisions which we have not heretofore mentioned, especially for phrases linking two predicates. These provisions have been made, but we shall not discuss the necessary syntactical maneuvers here. We shall limit ourselves to an explication of what is involved in the sub-process, "Find the Prime Mover." (Some discussion of the other two sub-processes may be found in an earlier paper by Abelson[1]).

The system takes the sentence C-P, constructs the predicate "controls C" and searches its memory for a belief of the form, "B controls C," where B is a concept more evaluatively consistent with P than was C. If none is found, then the sub-process has failed. If an appropriate B is found, the system then attempts to establish the credibility of the sentence B-P. If this attempt fails, there is a search for an alternative B. If this attempt succeeds, however, then a successful rationalization has been constructed which might be paraphrased, "The actor C was not really responsible for action P, but was merely the helpless pawn of the wishes of B." To give a more concrete example with foreign policy content, suppose that a system with a weak positive evaluation of Cambodia is presented with the sentence "Cambodia rejects U.S. advice." Suppose that this information is found credible (for example, in terms of higher-level subjective generalizations concerning the obstreperous behavior of Asian nations), and that the system subsequently attempts to rationalize the sentence. A successful outcome might well be the compound sentence, "Red China controls Cambodia's rejection of U.S. advice."

System Control of the Response Processes

The response processes outlined above are called upon by the system in an orderly sequence dependent upon the circumstances. This "calling sequence" constitutes a kind of executive function by which the system governs its operations. In the present system, the "executive" is prepared to respond only to assertions fed into the system in a standard sentence format. After each assertion has been dealt with, the system waits for another assertion. Because it does not operate autonomously in between externally triggered episodes, the present system is a "one-shot" rather than continuously functioning cognitive system. (In future versions, provision will be made for continuous functioning. The system will have a "Worries List" on which it stores various bothersome by-products of previous operations. When there is no input, the system will consult its Worries List to find something to think about, somewhat as Colby and Gilbert's[12] simulation of neurotic thinking runs iteratively through whole complexes of beliefs.)

The standard format for input sentences to the system is "S says C-P," where S denotes the source of the assertion. Specification of the source is crucial for the achievement of a realistic simulation. In order to be capable of evaluating what is said, the system must know who says it. Incoming assertions are potentially persuasive communications, and it is well known that the effects produced by persuasive communications are critically dependent upon the characteristics of the communicator (cf. Hovland, Janis, and Kelley[13]).

In order to use information about communication sources, the system must have prior knowledge (or beliefs) about what sorts of things have been said by whom. Fortunately, the syntax of our system readily permits such memory storage. Any known source, S, is simply

a concept in the system, and the construction, "says C-P" is a (compound) predicate. Thus constructions of the form "S says C-P" are (compound) sentences, potentially manipulable by the system in precisely the same ways as all other sentences. For example, "S says C-P" can be subjected to a Credibility Test and found credible because other persons sharing some of S's qualities are believed to have said things sharing some of the qualities of the unit (C-P). Or again, "S says C-P" can be entered into the Attempt Rationalization process and a rationalization found to the effect that S's statement is in the service of some ulterior purpose, or is understandable only as the view of some hidden figure controlling S, and so on.

The system "executive" decides which processes are to be applied in what order to "S says C-P" as well as which processes are to be applied to the inner sentence "C-P" itself. Of course the executive is not a homunculus; choices of processes are actually automatically made as a function of the system's prior evaluations of three elements of the input: the source S, the predicate P, and the compound concept (C-P). We cannot in this brief exposition give the flow chart for the entire executive program, but we can indicate the major branches.

The first thing the executive does is to test the credibility of the input components. Test failure gives rise to an "incredulity reaction," and processing stops. Test success, however, does not imply system acceptance of the input. The executive next examines its evaluation of S. If the source is negative, then the system is motivated to dismiss the source's assertion either by denial or rationalization, especially if the concept and predicate of the assertion are evaluatively inconsistent. If the source is positive, neutral or unfamiliar, then the executive next inspects the evaluative consistency of C and P.

The case of consistent evaluation sends the system quickly into a terminal branch. There is not much left for the system to do when a favorable source makes a credible statement that a good actor is engaged in a good action or a bad actor is engaged in a bad action, except to express joy or regret. (An improved version of the system might in the latter case attempt to construct a recommendation for dealing with the bad actor.)

The case of inconsistent evaluation is more complicated, and the two major sub-cases must be treated somewhat differently from each other. If a good actor is asserted to be engaged in a bad action, then the system is motivated primarily to deny this alleged fact or to deny that the assertion was made. If a bad actor is asserted to be engaged in a good action, the system is motivated primarily to rationalize this alleged fact or to find a reason why the assertion was made. At any rate, the executive has at its disposal all four of the above mechanisms, namely: Deny (C-P); Deny (S says C-P); Rationalize (C-P); Rationalize (S says C-P).

In one way or another (possibly including the failure of all processes attempted), system control finally passes from the executive, and a response is printed out. Altogether there are 32 possible types of responses in the present version of the system. These responses need merely be indexed symbolically so that the person running the simulation can find out what the system did. For entertainment value, however, the system has been programmed so that a prepared English statement denotes each response. Thus, particular exits from the system may be accompanied by lurid remarks such as, "That's just the kind of nonsense I'd expect to hear from him," or "That news shakes me up. Did he say anything else about it?," and so on.

When the denial or rationalization processes succeed, the supporting evidence or reasoning developed by the system is printed out. In order to do this, the system decodes its internal symbols into English by means of an internal "dictionary" provided to it, and then inserts these English words into appropriately pre-packaged linguistic frames such as "*(C)* only *(P)* in order to achieve *(R)*."

Thus the entire printed output of the system for each one-shot response is in more or less coherent English. We hope soon to be able to express the input in (simplified) English as well, so that a smooth "interview" with the system may be conducted. (At present, the input is symbolically coded. The programming language, IPL-V, in which the system is presently programmed renders the use of verbal input very clumsy.)

Effects of the Processes on the System

We have given many details of how the system produces responses of a resistive character, but have as yet said nothing about the changes which take place in the system itself. In the course of its machinations, the system undergoes four types of changes: 1) the creation of new instances to include in its vertical memory structure; 2) the gradual development of a "style" of rationalization for particular contents; 3) the storage of new beliefs found credible and not later denied or rationalized; 4) gradual evaluative changes.

The creation of new instances is an incidental by-product of the Credibility Test. In searching for an instance of a compound concept (C-P), the system may, for example, discover that an instance C_i of C forms an admissible compound concept (C_i-P) when paired with P. If (C_i-P) had not previously been stored as an instance of (C-P), it is so stored. Much more complicated (and thus more interesting) examples are within the range of system capability, but in a sense none of these newly created instances are very "creative," since they are all strongly implied in the original memory structure.

Development of a "rationalization style" is made possible by provision for the system to remember the way in which it has rationalized any particular sentence. When the need arises to rationalize a new sentence, the system first finds a sentence similar to the new one (by moving up and down the vertical network to "nearby" concepts and predicates), and then looks to see what rationalization sub-process and crucial element (B, Q, or R) may have been used on that similar sentence. If

such information can be found, a similar rationalization attempt is applied to the new sentence. Thus over the long run the use of a particular rationalization mode may snowball and give the system a certain stylistic character, whether it be paranoid (through excessive use of "Find the Prime Mover"), or apologetic ("Accidental By-product"), or Polyannic ("Reinterpret Final Goal"), or a blend of these.

The storage of new beliefs can come about at several points in the system. If an input unit C-P (or S says C-P) is found credible and not later denied or rationalized. it is stored when the system makes its output response. Moreover, the Credibility Test may discover and certify additional higher-level beliefs during the course of its examination of the qualities of C and P, and these too are stored when appropriate termination points are reached. Yet a third possibility is that certain new sentences will be generated by the application of the Credibility Test during a rationalization attempt. Success of the rationalization process automatically leads to storage of such new sentences. One restrictive feature embedded in the Credibility Test is worth emphasizing, however. Even though it is quite possible for both a sentence and its opposite to be found credible, in no case can a sentence C-P be found credible if the sentence C-not P is already stored in the system. Flat contradiction of prior beliefs, then, is not admissible in the present system. It is quite possible, however, for innocent-seeming beliefs to slip past the Iron Duke of resistance and later cause major shifts in emphasis of the belief system. The design of future versions of the system will take into account various other ways in which belief systems can ultimately be drastically modified (for example, by firm attachment of beliefs to particular sources who later become discredited, thus converting a belief into a "disbelief").

Evaluative changes are incorporated into the system in a rather novel way. We have postulated that changes in the evaluations of elements can occur only when the system is "thinking" about these elements. When two elements are simultaneously present "in thought," we postulate that "evaluative transfer" takes place between them. The system definition of what it means for a pair of elements to be simultaneously "in thought" is that a particular concept and a particular predicate are both under the direct control of the executive at a particular point in time. One clear occurrence of this condition is at the moment when a sentence C-P is input to the Credibility Test. At this moment (and at other comparable moments) a very small fraction of the evaluation of C is transferred (i.e., algebraically added) to the evaluation of P, and a very small fraction of the evaluation of P is transferred (algebraically added) to the evaluation of C. The effect of this transfer is that if C and P have the same evaluative sign, both evaluations will be boosted to greater quantitative extremity, but if C and P have opposite evaluative signs, both evaluations will tend to "run down" toward neutrality. It is precisely the latter effect which we take to be basically disturbing to the system, thus motivating the use of the denial and rationalization processes. But if these resistive attempts consume a great deal of cognitive effort and ultimately fail, the system will be repeatedly and inescapably exposed to small doses of evaluative transfer which may aggregatively result in substantial evaluative change, including changes of evaluative sign. This will especially be the case when one or the other element has a weak initial evaluation. A very "tight" or "closed" belief system will tend not to be subject to this effect because it will "have an answer for everything," but a more "open" system with many internally unresolved questions and qualified beliefs will be more vulnerable to appropriate persuasion, as indeed one would expect.

These common sense outcomes are inherent in the design of the system, but we have not had sufficient experience in simulating real belief systems to know whether the long-run behavior of our computer model will be fully realistic in most major respects.

Putting a Memory Structure into the System

Readers who have dutifully followed our presentation in all its sundry details may have long since wondered how the system gets its memory structure in the first place. The basic answer to this question is that the simulation user must put it in himself, albeit there are at least three ways in which this can be done.

One method is to invent a structure *ad lib*. If this (or any other) method is to be psychologically interesting, one must have real subjects whose responses can be compared with the responses of the simulated system. One might possibly have real subjects *learn* the invented structure. The content would have to be fairly limited in scope and lacking in prior relevance to the subjects—a dramatic encounter on Planet X9 between weird good and bad creatures, for example. This method is not without its problems, but serious consideration is being given to the possibility of employing it for purposes of testing the model.

A second method is to *interview* a real subject to elicit his own belief structure. This procedure is very laborious but feasible provided the subject is articulate and cooperative. The interviewer starts with a general concept defining the topic under consideration and systematically develops the network of further concepts and predicates by means of eliciting frames such as "What are some examples of (concept)?" "What actions do you associate with (concept)?" etc. This procedure is quite similar in spirit to those of modern ethnographers such as Frake,[14] who explore the beliefs held collectively by members of a given culture. We have not used this procedure in full elaboration, but several interesting opportunities to do so are being considered.

The third course is the one presently being undertaken by the first author. A person is chosen whose beliefs are well understood by the investigator, and these beliefs are paraphrased in order to fit them into the memory structure. The identity of one individual so chosen has previously been indicated. His belief system is notably

"closed," and the simulation of this system will provide a good intuitive test of certain features of the model. The results will be reported elsewhere. The simulation of the belief systems of other individuals with very different views is also being contemplated, but this step cannot be undertaken lightly, since the paraphrasing procedure is extremely difficult. One might suppose that fully automatic content analysis methods could be applied to the writings and speeches of public figures, but there is an annoying technical problem which renders this possibility a vain hope, even presuming that other extreme technical difficulties could be resolved. This annoying problem pertains to the implicit "knowledges" necessary to define the full belief structure. The prose of public figures ordinarily omits many obvious facts and definitions presumably taken for granted both by him and by his audience.

The three procedures outlined above all concern empirical tests of the full simulation model. More limited partial testing is also possible—indeed, much more feasible.

A Preliminary Empirical Study

One portion of the system eminently amenable to experimental study is the Credibility Test. Since the operation of the test is independent of system evaluations of the elements involved, one can employ innocent or nonsense content in setting up an empirical analogue of the test situation. This was in fact done in a study by Gilson and Abelson.[15]

The study was designed to test a somewhat puzzling implication of the success rule for the sub-process where the system searches for *instances* of C and P. That rule, it will be remembered, is that C-P is found credible when at least half of the instances of C are connected to at least one instance of P. There is an asymmetry in this rule, the full consequences of which were not apparent when the rule was initially invented.

The point may be appreciated by consideration of the following two simple cases:

1) There are altogether three instances of a concept, only one of which is connected to a given predicate;

2) There are altogether three instances of a predicate, only one of which is connected to a given concept.

These two cases are entirely equivalent except for the interchange of roles of concept and predicate. Yet the success rule is only satisfied for the second case, not for the first. The asymmetry boils down to the condition that the system is willing to take one predicate instance out of several as roughly representative of the whole predicate class, but is not willing to take one concept out of several as roughly representative of the whole concept class. That is, the system is more willing to generalize over a set of related predicates than over a set of related concepts.

In the Gilson and Abelson study, subjects were variously presented with "evidence" expressed in the two simple forms above, and queried whether they would agree or disagree with the corresponding inductive generalizations. Examples of questions using the two respective forms are as follows:

1. Altogether there are three kinds of tribes—Southern, Northern, Central.
   Southern tribes do not have sports magazines.
   Northern tribes have sports magazines.
   Central tribes do not have sports magazines.
   Do tribes have sports magazines?
2. Altogether there are three kinds of magazines—sports, news, fashion.
   Southern tribes do not have sports magazines.
   Southern tribes do not have news magazines.
   Southern tribes have fashion magazines.
   Do Southern tribes have magazines?

A variety of subjects, objects, and verbs were used to explore the effects of specific sentence components on the tendency to accept generalizations of these forms.

Although the difference was not of substantial magnitude, there was very clearly a greater tendency to agree with questions of the second form than the first form. In other words, the subjects were more willing to generalize over predicates than over concepts, as predicted. Another finding of the study was the rather wide variation in agreement frequency as a function of the sentence verb. The verb "have," for example, was much more conducive to the acceptance of generalizations than the verb "like."

Our simulation model does not presently take effects of this latter kind into account, but as the accumulation of experimental results improves our understanding of the complexities of subjective logic and resistive mechan. s much improved versions will be possible. Obviously we have barely begun to scratch the surface in conducting relevant empirical studies, but it is hoped that by converging on the study of belief systems from many angles at once, substantial progress in this fascinating area will be made in the next few years.

References

1 Abelson, R. P. and Rosenberg, M. J. "Symbolic Psycho-logic: A Model of Attitudinal Cognition," *Behav. Sci.*, 3 (1958), 1-13.
2 Abelson, R. P. "Modes of Resolution of Belief Dilemmas," *J. Conflict Resolution*, 3 (1959), 343-52.
3 Rosenberg, M. J. and Abelson, R. P. "An Analysis of Cognitive Balancing" in M. J. Rosenberg *et al., Attitude Organization and Change*. New Haven: Yale Press, 1960.
4 Abelson, R. P. "Computer Simulation of 'Hot' Cognition" in S. Tomkins and S. Messick, eds., *Computer Simulation of Personality*. New York: Wiley, 1963.
5 Heider, F. "Attitudes and Cognitive Organization," *J. Psych.*, 21, 1946, 107-12.
6 Heider, F. *The Psychology of Interpersonal Relations*. New York: Wiley, 1958.
7 Festinger, L. *A Theory of Cognitive Dissonance*. Evanston, Ill.: Row, Peterson, 1957.
8 Osgood, C. E. and Tannenbaum, P. H. "The Principle of Congruity in the Prediction of Attitude Change," *Psych. Rev.*, 62, 1955, 42-55.
9 Rokeach, M. *The Open and Closed Mind*. New York: Basic Books, 1960.
10 Smith, M. B., Bruner, J. S., and White, R. W. *Opinions and Personality*. New York: Wiley, 1956.
11 Miller, G. A. "Language and Psychology" in E. H. Lenneberg (ed.), *New Directions in the Study of Language*. Cambridge, Mass.: M.I.T. Press, 1964.
12 Colby, K. M. and Gilbert, J. P. "Programming a Computer Model of Neurosis," *J. Math. Psych.*, 1 (1964), 405-17.
13 Hovland, C. I., Janis, I. L., and Kelley, H. H. *Communication and Persuasion*. New Haven: Yale Press, 1953.
14 Frake, C. O. "The Ethnographic Study of Cognitive Systems" in T. Gladwin and W. C. Sturtevant, eds., *Anthropology and Human Behavior*. Washington, D. C.: Anthropological Society of America, 1962.
15 Gilson, C. and Abelson, R. P. "The Subjective Use of Inductive Evidence," *J. Pers. Soc. Psych.*, in press, 1965.