STANFORD ARTIFICIAL INTELLIGENCE LABORATORY
MEMO AIM-253

STAN-CS-74-471

THE INTERACTION OF INFERENCES, AFFECTS, AND INTENTIONS, IN A MODEL OF PARANOIA

by

Bill Faught
Kenneth Mark Colby
Roger Parkison

SUPPORTED BY
NATIONAL INSTITUTE OF MENTAL HEALTH
and
ADVANCED RESEARCH PROJECTS AGENCY
ARPA ORDER NO. 2494

DECEMBER 1974

COMPUTER SCIENCE DEPARTMENT
School of Humanities and Sciences
STANFORD UNIVERSITY

STANFORD ARTIFICIAL INTELLIGENCE LABORATORY
MEMO-AIM253

DECEMBER 1934

COMPUTER SCIENCE DEPARTMENT
REPORT NO. STAN-CS-74-471

THE INTERACTION OF INFERENCES, AFFECTS, AND INTENTIONS IN A MODEL OF PARANOIA

Bill Faught (1)
Kenneth Mark Colby (2)
Roger C. Parkison (3)

Abstract:

The analysis of natural language input into its underlying semantic content is but one of the tasks necessary for a system (human or non-human) to use natural language. Responding to natural language input requires performing a number of tasks: I) deriving facts about the input and the situation in which it was spoken; 2) attending to the system's needs, desires, and interests; 3) choosing intentions to fulfill these interests; 4) deriving and executing actions from these intentions. We describe a series of processes in a model of paranoia which performs these tasks. We also describe the modifications made by the paranoid processes to the normal processes. A computer program has been constructed to test this theory.

(1) Research Assistant, Department of Computer Science, Stanford University
(2) Research Professor of Psychiatry, University of California at Los Angeles
Formerly Adjunct Professor of Computer Science, Stanford University
(3) Research Assistant, Department of Computer Science, Stanford University

This research is supported in part by Grant PHS MH 06645-13 from the National Institute of Mental Health, in part by Research Scientist Award (No. I-K05-K-14,333) from the National Institute of Mental Health to the second author, and in part by the Advanced Research Projects Agency of the Department of Defense under Contract DAHC15-73-C-0435.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the National Institute of Mental Health, the Advanced Research Projects Agency or the U. S. Government.

THE INTERACTION OF INFERENCES, AFFECTS, AND INTENTIONS IN A MODEL OF PARANOIA

Bill Faught
Kenneth Mark Colby
Roger C. Parkison

INTRODUCTION

The analysis of natural language input into its underlying semantic content is but one of the tasks necessary for a system (human or non-human) to use natural language. Responding to natural language input requires performing a number of tasks: 1) deriving facts about the input and the situation in which it was spoken; 2) attending to the system's needs, desires, and interests; 3) choosing intentions to fulfill these interests; 4) deriving and executing actions from these intentions.

We shall describe a series of processes which respond to natural language input. The development and implementation of these processes in a computer program are part of an on-going research project to construct a simulation of a person whose behavior is dominated by the paranoid mode of thought processing. The model consists of 2 major parts -- 1) recognizing processes which use a sequence of pattern-matching rules to recognize English input expressions; and 2) response processes which include: a belief structure, a process to make inferences that alter the belief structure, a process which attends to affects, a process which derives intentions, and a process which calculates actions to perform. We shall describe only the response processes here. Details of the operation of the language recognizer are contained in a previous communication [Colby, Parkison, and Faught, 19741.

According to the theory embodied in the simulation model [Colby, 1975], the paranoid mode influences certain types of beliefs, inferences, affects, intentions, and actions. In the paranoid mode, a person scans natural language input, and the inferences from that input, for evidence which judges an action, desire, or state of the self to reflect an inadequacy or defectiveness of the self. Upon finding such evidence, an attempt is made at simulating acknowledgement of this inadequacy. If belief in inadequacy were accepted or acknowledged as true, humiliation would result. The detection of impending humiliation in the simulation serves as a warning not to execute the acknowledging procedure. Instead an alternative simulation is attempted in which wrongdoing is attributed to others. Since no warning signal of humiliation results, the procedure for blaming others is executed. The outcome of this alternate strategy is 1) to repudiate that the self is to blame for inadequacy, and 2) to ascribe blame to other human agents. This transfer of blame is reflected in the ongoing linguistic behavior of the paranoid patient in a psychiatric interview.

Our simulation is a vehicle for testing this theory. As such it must contain simulations of all the normal processes which the paranoid mode influences. These processes include: a system of beliefs about the world, about the current situation, and about the self; a mechanism for making inferences to derive evidence for new beliefs; a set of affects whose states reflect current needs, desires, and interests; an affect mechanism to modify affects and reflect their states in modifications to other processes; a set of intentions defining goals which can satisfy needs and desires; and a mechanism which derives these intentions and computes and executes actions to satisfy their goals. In addition, the simulation must contain a specific embodiment of the abnormal processes of the paranoid mode.

Participating in a psychiatric interview is an ideal task for a simulation of the processes which are included in the theory, because the task requires so many of the

cognitive, affective, and conative processes that are used by the paranoid mode. When participating in such an interview, a person (particularly if he has already participated previously in such interviews) brings with him many preconceived ideas about the purpose of such situations and what happens in them. He is motivated by his self-interests and has goals specifying what he wants to accomplish during the interview. He has plans about how he intends to obtain those goals. He has expectations about how the interviewer will respond to his plans carried out in actions. He has a cognitive ability to observe and evaluate the actions that are taking place to determine whether his goals are being achieved or whether some new situation arises with which he must cope. He has needs and desires which are tied to the success or failure of his actions as well as to the interaction of the participants in the situation. Finally he has the ability to posit new goals and derive courses of action to obtain those goals, possibly altering the course of the situation in which he is participating, while he is in the middle of that situation, As he is attempting to steer the situation, he is again perceiving the ongoing situation and measuring the success of his actions.

Previous simulations involving inferences, affects, and intentions have tended to be limited to one of these processes. Belief system programs have concentrated on making credibility judgements, building belief structures, and answering questions. Affect simulations have been divorced from the influence that affects have on subsequent behavior, Simulations having goals and intentions leading to actions have tended to be devoid of affect and self-interest, and usually have been focused on abstract problem solving, rather than on deriving actions for self-interest intentions.

Because the paranoid mode influences all of these processes, it was necessary to define and implement all of them.

The three major response processes in our simulation deal with inferences, affects,

and intentions. These normal processes are the background for a theory of the paranoid mode. We have placed emphasis on their simplicity, clarity, and separation from each other, and separation from the modifications made by the abnormal paranoid mode. By emphasizing the separation between them we provide a clear base for applying the various parts of the theory, as well as the opportunity to improve and expand each process. Previous versions of the model of paranoia [Colby, Weber, and Hilf, 1971] had no separation of normal and abnormal processes, making changes in the implementation of the theory or changes in the underlying processes very difficult.

The interview situation constrains the processes in several ways. There are only a few sets of facts which the model (known as PARRY) needs to make inferences about: the interviewer's actions, the model's own immediately preceding actions, the course of the interview so far, and predictions about the future course of the interview. Only three of eight primary affects [Tomkins, 1962] are simulated: anger, fear, and shame; the others we consider to be of secondary importance in simulating a personality dominated by the paranoid mode. There is a limited set of intentions and subsequent actions that a person may expect to use in an interview. Actions of the model are limited to linguistic behavior in an interview. Intentions must therefore be limited to obtaining some situation through linguistic performance. Finally, the interviewer is also limited to linguistic behavior and therefore to the same type of goals.

OVERVIEW

An input expression typed by the interviewer is first recognized by the pattern-matching module, and then processed by the response module which we are about to describe. The **recognizer** (previously referenced) uses a sequence of pattern-matching rules to

transform the input expression into a reference to its semantic content in memory. Within the responder are five processes which react to the input based on the contents of the input, the state of the belief system, current needs and desires, and the goals and expectations of the model. These five processes are:

INTERFACE -- This routine is responsible for extracting and categorizing all the input from the recognizer. The English input expression is scanned for style information not contained in the recognizer's results. Global anaphoric references are resolved according to an expectancy list. The information relevant to responding to this input is retrieved-from memory.

INFERENCE -- Using the semantic content of the input accessed by INTERFACE and the current state of the belief system, this procedure makes inferences about the current state of the world, particularly about the interviewer, his beliefs, his intentions, about the interview, and about the personal interaction between the model and the interviewer. Inferences may add evidence for a belief being true, or may conclude positively, without a doubt, that a belief is true. The process of adding evidence to a belief's truth may directly influence affects.

AFFECT -- Using the current state of the three primary affects (anger, fear, shame) and the new information derived by INFERENCE, this routine computes new levels of affects. The affects are the primary motivation for determining the response to be given.

INTENTION -- This routine maps affects, beliefs, the current input, and previous intentions into the current intention. Competing intentions are resolved according to their intrinsic importance. From this intention an appropriate action is computed which will attempt to satisfy the goal of the intention (e.g., to reduce negative affects).

REPLY -- The desired action is performed by locating and expressing the

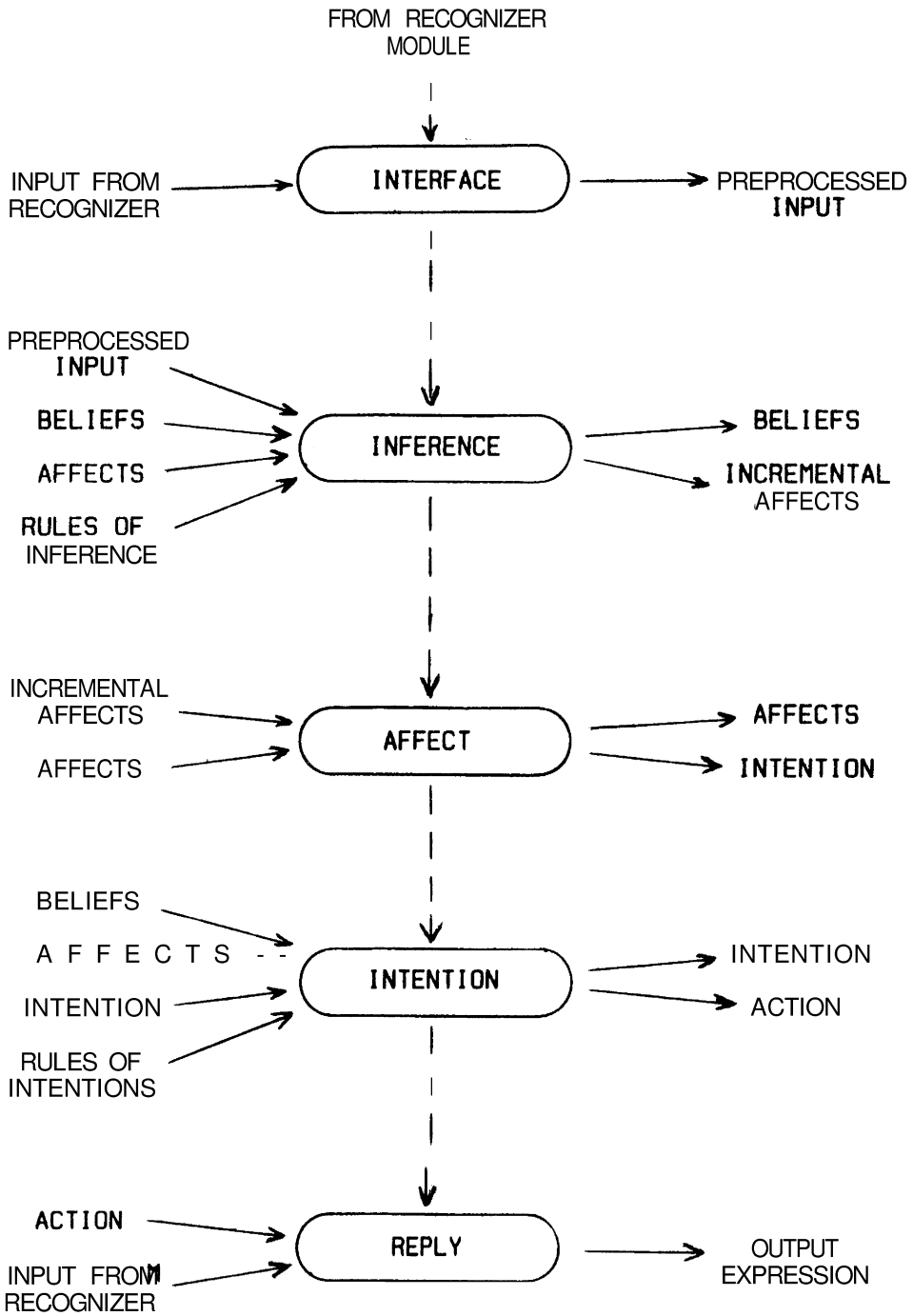

Figure 1.

Overall Flow Diagram of Response Processes

Flow diagram of data (solid arrows) and process control (dotted arrows) in the model.

proper English output sentence. An expectancy list of potential anaphoric references for the next input is set.

Each of these processes is applied to each input expression from the interviewer, although some may be more or less important for a particular input. Our aim has been to get as far away as possible from a question-answering or stimulus-response paradigm which has characterized many cognitive simulations (including previous versions of PARRY) by centering the origin of the motivation for each response in the AFFECT process, rather than in the input expression.

We shall describe each process in order, reserving the description of modifications made by the paranoid mode for a later section.

INTERFACE

When the recognizer processes an input expression, it provides results as follows:

1) a pointer to a data structure in memory containing information about the semantic content of this input,

2) a list of the English input with the recognizer's dictionary translation for each input word, and

3) parameters indicating the number of misspelled and unrecognized words,

INTERFACE completes the extraction of information useful to the responder from the input. INTERFACE has little in common with the other processes theoretically, and serves mostly to enhance the implementation of the model.

INTERFACE first adds to this information by examining the English input expression and its canonical translation to extract characteristics of the interviewer's style. The level of confidence in the recognizer's input is determined based on the number of

symbols in the input expression which could not be recognized as words, and the number of spelling mistakes which could be corrected. The amicability of the input is determined according to the presence or absence of positive- ("good", "nice", "please") and negative- (swearwords, "lousy", "peculiar") affect words. Because such words are often used as spurious modifiers ("How do you like the lousy hospital?"), they are usually not recognized as an important part of the semantic content of the input by the recognizer, especially in factual questions. These affect words are detected by the presence of their dictionary translations in a list constructed by the recognizer. From the presence of strong affect words, a general measurement of the interviewer's positive or negative affects is made. Style information is used in the INFERENCE process as antecedents of theorems. This completes the extraction of information from the input expression.

The semantic content of the input may indicate that the input represents one of several anaphoric references. The following anaphora are used in our simulation:

Input Anaphora		Examples of English expressions which the recognizer rewrites to the anaphora on the left
GO_ON	←←	"Tell me more, what happened"
ELAB	←←	"Explain, what do you mean"
WHY	←←	"Why, for what reason, how come"
HOW	←←	"How, in what manner"
WHEN	←←	"When, when does that happen"
WHERE	←←	"Where, where does that happen"
WHO	←←	"Who, who did that"
WHAT	←←	"What, for what"
YOU_DO	←←	"Do you, what do you do"
THEY_DO	←←	"Do they, what do they do"
HOW-KNOW	←←	"How do you know"
HOW-MUCH	←←	"How much, how many"
HOW-LONG	←←	"How long"

These anaphora represent genera1 references which may apply to previous output expressions.

If the input is one of these anaphora, it is looked up on an expectancy list of anaphoric references, which has been established by previous output expressions. A new pointer to a data structure containing the complete semantic content of the input is then derived. For example, the output statement:

"I AM IN THE PALO ALTO VA HOSPITAL"

would set up an expectancy list similar to the one below which would cause the anaphora on the left to derive the same pointer as the complete expression on the right:

GO_ON	t t	"Tell me more about the hospital"
WHERE	t t	"Where is the hospital"
WHY	t t	"Why are you in the hospital"
HOW-LONG	←←	'How long have you been in the hospital'

The new pointer is then used in the same manner as a non-anaphoric input.

Local anaphora (he, him, his, she, her, hers, it, its, they, them, their) are detected by the recognizer during its processing. The recognizer uses the expectancy list to fill in the intended concept before using its pattern-matching rules to recognize the input.

INTERFACE also uses the pointer to retrieve the semantic information for this input from the memory. Such information includes: the type of input (e.g., threat, insult, apology, factual question), the topic under discussion, a pointer to factual answers if this is a factual question, and a pointer to inferences in the memory which will be used based on this input.

INFERENCE

The INFERENCE process is used for several tasks which are important to the model PARRY.

One of PARRY's goals is to get help. In order to best seek help, the model must be

able to make inferences about the interviewer -- the interviewer's abilities, interests, intentions, traits, and probable future actions. The model must also evaluate the progress of the interview and the success of attaining its goals so that it may better direct itself in its pursuits.

Secondly, the model must beware of threatening situations, arising from either mental or physical threats. For this it must compare the interviewer and the interview to its concept of typical behavior, and judge them as normal or abnormal, and therefore predictable and non-threatening or unpredictable and potentially threatening.

Finally, in obtaining the help of a psychiatrist, the model is sometimes called on to examine and evaluate its own behavior. For this task it must examine its own behavior in the interview and evaluate the psychiatrist's opinion before making a judgement about its own behavior.

For these tasks we constructed an inference process which works with a belief system. The belief system contains approximately 50 beliefs. [See Appendix 2.] Beliefs refer only to topic areas in which there can be evidence to change the belief in the course of one interview. Such areas refer to the interviewer, the interview, and the current state and intentions of the self, and not to relatively unchanging facts about the world, or its own past history. Beliefs begin with default values indicating the self's assumptions prior to the interview, and change during the course of the interview.

A belief is represented in the model by a data structure with the property of TRUTH whose value corresponds to the amount of evidence that the belief is true. The truth value ranges from 0 to 10, 0 indicating no information, 10 indicating enough evidence to conclude that the belief is true. INFERENCE can conclude that a belief is true (truth value set equal to 10) or add to a belief's truth value (truth value incremented by a positive integer).

The negation of a belief is an entirely separate belief with a truth value from 0 to 10. For example, the following four beliefs are in the system:

DDHELP -- the interviewer desires to help PARRY
*DDHELP -- the interviewer does not desire to help PARRY
DDHARM -- the interviewer desires to harm PARRY
*DDHARM -- the interviewer does not desire to harm PARRY

Note the usefulness of having all four beliefs specifically with their own truth values, permitting us to conclude as true the belief *DDHELP quite independently of the belief DDHARM. Also we allow competing evidence to accumulate for both DDHELP and *DDHELP without having evidence for one cancel evidence for the other. Contradictions may arise if enough evidence accumulates to infer both the belief and its opposite. The current model notes such contradictions, but no action is taken.

Beliefs also have a property AR (Antecedent Rules) which lists all of the names of the rules of inference in which the belief is an antecedent. This makes possible the easy access of the appropriate rules of inference should a particular belief be found to be true.

The INFERENCE process works with approximately 150 rules of inference. [See Appendix 31. Inferences correspond to the ability to draw new conclusions about new situations. In our simulation, inferences make possible the evaluation of the interviewer and the interview, examination of the self's actions, and prediction as to the future. behavior of the interviewer. They invoke no actions, intentions, or affects themselves; instead their results are used as data by all other processes in the simulation for making cognitive evaluations about the world.

A rule of inference in the model is represented by a node which has a name, a list of antecedents, and a consequent. The antecedents are each tested for truth, and a logical "and" applied to the resulting set of logical values from the antecedents. An antecedent may be:

1) a belief, which is true if its truth value is true, 2) a negation of a belief, in which case the antecedent is true if the truth value of the belief is not yet true (i.e., not yet equal to 10, the maximum), 3) a pointer to the semantic content of an input expression, which antecedent is true if the current input is this expression, or 4) an arbitrary function which is evaluated, usually testing a measurement of the interview or an affect strength. A consequent may be either a belief with a truth value of 10, in which case that belief is concluded as true, or a belief with an incremental truth value, in which case the increment is added to the truth value for that belief.

The INFERENCE process begins by gathering' from various processes, evidence that new inferences can be made. The input semantic content is examined for inferences which can be made based on this input having been expressed. Measurements are made of the interview concerning length, variation, dominance of the participants, and politeness (e.g., the number of input expressions, the number of sentence anaphora used, the number of new topics introduced by the interviewer). The affects are examined for inferences which can be drawn from the state of the affects. Whenever a rule of inference is found which is a candidate for INFERENCE (i.e., one of the antecedents of the inference has just become true) the name of the rule of inference is put on a list of candidates INFERLIST. [INFERLIST typically contains IO-20 candidate inferences.] A process called INFER is invoked which tries to infer each candidate inference on INFERLIST in turn until the list is exhausted. INFER 1) checks to see that the consequent is not already true, 2) determines the truth values of the antecedents, and 3) performs a logical "and" on the resulting logical values. Whenever an inference is made and a new belief concluded to be true, the rules of inference for which the new belief is an antecedent become candidates and they are added to INFERLIST to be tried (using the property AR described above). Thus the INFERENCE process makes all possible inferences based on the new information it has whenever it is invoked.

The end result of the INFERENCE process is to update the state of the belief system to reflect new information obtained from the input,

AFFECT

In terms of the very existence and motivation of the model, AFFECT is the fundamental process in the simulation. AFFECT reflects the overall state of the system and processes on that state which provide motivation for all other action in the system. AFFECT spurs INTENTION to find a goal which will satisfy affect-requirements; AFFECT uses INFERENCE to interpret events occurring in the situation to determine if affect-requirements can be satisfied. In the current model these requirements involve reduction or stabilization of negative affect.

Each affect in the simulation represents a type of motivation for one aspect of the system and a desire or need for that aspect, When these needs or desires become activated, either through self-motivation or from perception of external events, they invoke a sequence of actions to attend to the need. Such actions include: perceiving whether or not the current situation in which the self is participating is in any way congruent with the current affects (or whether participation in the current situation should be terminated), determining a goal which will satisfy the need, determining and executing actions to satisfy the goal, and perceiving the effects of the actions to judge whether the desired goal was achieved. Further, AFFECT may influence the execution of other processes in situations of high affects. AFFECT may override INFERENCE and not allow certain inferences to be made (with significant side effects), or allow certain inferences to be made with less evidence than is normally required. Also, AFFECT may override the normal determination of an intent by INTENTION if one affect needs immediate attention. Thus in all of these effects, the affects

and the process AFFECT in our simulation serve the purpose of relating cognitive ability to the overall system; i.e., the desires and needs of the overall system are embodied in the states of the affects.

The AFFECT process consists of three primary affects, rules for interpreting information gathered elsewhere in the simulation which influence the affects, rules for modifying the affects, and rules for influencing other processes based upon the state of the affects, The three primary affects represented are anger, fear, and shame. High degrees of anger and fear are symptomatic of the paranoid mode. Intense shame is the primary causative affect of the paranoid mode, according to the theory embodied in the model. [These affects are also normal -- it is their intensity and duration that is abnormal in the paranoid mode.] Affects are represented as variables which may take values from 0 to 20, 0 being low, 20 high.

AFFECT gathers information for modifying affect states from two sources: the cognitive process INFERENCE and the conative process INTENTION. Certain beliefs carry with them the ability to produce affect changes when evidence is found that these beliefs are true. As INFERENCE changes a belief's truth value or concludes a belief to be true, AFFECT is noting which affects are to be modified and in what manner. The size of change in the affects depends both upon which belief there is evidence for and the amount of evidence there is that the belief is true: If there is evidence causing the belief to be concluded, then the size of the change is twice as much as if the belief were only to be added to. This information is stored in three incremental variables (AJUMP, FJUMP, and SJUMP) indicating how much to raise or lower each affect. For example, if evidence is found supporting the belief DDHARM (the interviewer desires to harm PARRY), the incremental variable FJUMP (indicating the raise- in fear) is set to a positive (non-zero) value. Information

about affect changes is also gathered from INTENTION. When certain actions are performed (e.g., attacking the interviewer, withdrawing), . the performance itself may modify affects by reducing them. This modification is performed by INTENTION after AFFECT is completed.

When INFERENCE is complete, the three affect variables and three incremental variables are used to determine new affect values. In a non-paranoid mode each affect is independent; the new value is purely a function of the old value and the increment. The values of the incremental variables range from 0 to 1, 0 being no change in affect level, 1 incrementing the affect to 20, its highest value, an intermediate value (e.g., 1/2) raising the affect level an amount-proportional to the incremental variable (e.g., 1/2 the difference between the current value and the maximum value possible). The values of the affects asymptotically approach 20 so that early increments have the most impact. In a later discussion of the paranoid mode we will see a case where the affects are not independent.

The result of updating the affects may be that one of the affects is extreme and requires immediate attention. AFFECT first decides, using beliefs from INFERENCE, if the current situation is appropriate for trying to satisfy a need of the system, or whether the situation should be changed. (In our simulation this involves ending the interview). If the situation is still appropriate and an affect has an extreme value, AFFECT triggers the proper intention so that this affect requirement is met immediately, outside the normal intention . selection process. The two intentions selected in this manner are called PSTRONGFEEL and PPARANOIA. Otherwise AFFECT invokes INTENTION, which may access the affect states in its processing.

Situations in which AFFECT overrides the other processes will be seen later in the description of the paranoid mode.

INTENTION

In the act of responding to an outside event, a complex system will take into account its knowledge of the world and of the current situation, its own needs and desires, and its previous experience with similar situations. It uses all this input and fashions some directed sequence of actions tending toward a goal. INTENTION has this function in our simulation. INTENTION is the process which gives substance to needs and desires by establishing goals for subsequent actions.

INTENTION examines current affects, beliefs, and previous intentions to formulate new intentions with their associated goals. The result may be several competing viable intentions. One is selected as being most important (according to a predetermined ordering) and it becomes the current INTENT. An action which will carry out this intention is then computed, based on the input, the beliefs, the current affects, and the intention.

There are 12 intentions represented in our simulation. [See Appendix 4]. Intentions are represented in a similar manner as beliefs -- a data structure with a property STRENGTH indicating its current strength. STRENGTH ranges from 0 to 10. Intentions differ from beliefs in that an intention becomes viable when its strength crosses a threshhold at 5, and the property STRENGTH of intentions can be both incremented and decremented. The strength of an intention is modified by a mechanism similar to the INFERENCE process, Changes in truth values of beliefs and changes in affects are tested by a set of rules, which modify the intention strength. These rules are similar to rules of inference in that they have antecedents and consequents. They differ from rules of inference due to the nature of intentions which are consequents: the rules may be used repeatedly for the same consequent intention. These rules may also examine a previous action or the result of a

previous action, thereby testing whether a previous intention was successful. When all of the intention values have been updated, the most important intention is selected according to a scheme which allows several intentions to compete with each other for satisfaction. First a subset of all intentions is determined consisting of those intentions whose values are greater than an activation threshhold. From these active intentions, one is selected as the current INTENT according to a predetermined ordering. For example the model starts an interview with the strength of the intention PINTERACT equal to 5. Since it is the only intention activated at that time, it is chosen as the current INTENT. Later in the interview, if the intention PHELP is activated, it is then chosen over PINTERACT as the current INTENT. [See Appendix 4 for the ordering of the intentions.1

To compute an action, each intention has a program which maps the intention, affects, beliefs, and the input into an action. Actions in our simulation take a linguistic form -- the only actions allowed are those which can be accomplished by means of a teletyped output. Examples of such actions are insults, lying, withdrawal, probing for information, praise, changing the subject, and factual answers to factual questions. Actions may themselves include instructions to modify affects, just on the fact of having performed a particular action. The performance of a particular action may also modify an intent, resulting in its goal being satisfied or modified. The specific action to be performed is sent to REPLY for execution.

For example, the intention PMAFIA has the goal of commenting on the doctor's Mafia connections. (The intention was probably invoked by the interviewer mentioning the Mafia before PARRY mentioned anything about gangsters or criminal activity). The program for PMAFIA first checks to see if the affect of fear is already high; if so, the action is to panic (because there really is a good reason to fear the interviewer). If fear is low and if it has

already been proved that the interviewer plays games, then the action is to question why the interviewer plays games and to lower the anger caused by the game-playing. Otherwise, the action is to probe the interviewer for the reason why he brought up the topic of the Mafia. In any case, the intention PMAFIA is decremented by a small amount so that it will not be activated for the next input.

REPLY

The function of generating an English language expression from a specific action is performed by REPLY. The input to REPLY is the action derived by INTENTION and the pointer to the semantic content of the input. For questions about demographic data about PARRY, the action is often just to answer the question. In this case the pointer to the input semantic content is used to extract a reply based on the factual data in the memory. Otherwise, the name of the action is used to locate the appropriate English output.

The output is chosen from a list of English expressions involving the same concept. (We are currently working on the problem of language generation with the intent of making a generation process which is in line with the remainder of the system). REPLY performs a number of bookkeeping functions such as keeping record of what has been said, finding an alternate output when a set of expressions has been exhausted, and setting up global anaphora on the expectancy list for the next input. REPLY also scans the model's own output for concepts which may modify its affects because it mentioned them. (For example, the model becomes somewhat fearful when it introduces its story about its argument with the bookie for the first time.) REPLY also reduces affect strengths by a time-decay factor, corresponding to the natural weakening of affect strength over time.

The end result of REPLY is an English output expression which is typed to the interviewer on a teletype or other display device.

THE PARANOID MODE

In the preceding sections we have described the processes and how they work when the simulation is in a normal (i.e., non-paranoid) mode. The paranoid mode represents an abnormal mode of processing, one which interrupts and overrides the normal processing by substituting some of its own subprocesses. The paranoid mode has its greatest effect on the two processes of INFERENCE and AFFECT.

According to the theory, the input and the inferences from that input are scanned for reference to evidence of an inadequacy or defectiveness of the self. In the model, such evidence is detected when evidence for one of a number of self-humiliation beliefs is found. These are beliefs about the self for which a large degree of humiliation is evoked if evidence is found that they are true (humiliation being extremely "painful" and to be avoided). In PARRY there are four of these beliefs: self is dishonest, self is stupid, self is crazy, and self is worthless. (See Appendix 2 for these core beliefs and others which represent their semantic equivalents.) A number of inferences draw conclusions which add evidence to support these beliefs, e.g. the interviewer believes PARRY is crazy, the interviewer believes that PARRY cheated someone, the interviewer believes that PARRY is not understanding his questions. Note that the type of new information which will infer these beliefs is attribution from others about the self's indequacies. It is assumed that the self has come to some sort of equilibrium of inferences after previously having thought about its own actions.

When one of the four self-humiliation beliefs is supported by evidence from an inference, INFERENCE sets the incremental affect variable for shame to a positive value, which is then detected by AFFECT. AFFECT uses the value to raise shame. If shame crosses a theshhold for paranoia, the paranoid mode is activated.

The first effect of the paranoid mode is to reject the belief which led to the increase in shame by resetting the truth value of the belief to its previous value. Instead, an alternate belief is inferred which would explain the interviewer's belief about the self; that alternate belief is generally that there is something wrong with, and/or malevolent about, the interviewer (e.g., "The interviewer must be incompetent if he thinks I'm crazy"). The next effect of the paranoid mode is to reduce the shame level now that the offending belief has been rejected. The shame level cannot be completely reduced to its previous state due to the severely disruptive processes that have been activated, so it is reduced by a factor proportional to its current value. -The third effect is to mitigate all of these disruptive processes by establishing an intention (called PPARANOIA) which results in a strong action, usually using the alternate belief just posited. Typical actions are: attack, insult, withdraw, lie. This intention (PPARANOIA) of responding to and possibly expressing the strong affects activated by the current input takes top priority over all the other intentions.

The paranoid mode strongly influences all three affects in the model. Each affect has a base value which it cannot fall below. These three affect bases (ANGERO, FEARO, SHAMEO) are 0 at the start of the interview. When the paranoid mode is activated, several affect changes occur: 1) The affect bases of anger and fear are raised an amount proportional to the shame affect, resulting in residual anger and fear after the paranoid mode is deactivated. 2) The base for the shame affect is raised to one half the highest level obtained by shame, with the result that the paranoid mode becomes easier to reactivate as the interview progresses. 3) The influence of all three incremental affect variables is enhanced when the shame affect level is high, resulting in more volatile fear and anger.

Once the paranoid mode is activated, it remains activated until the shame affect drops below the threshhold of paranoia (due to the time-decay of affects). However, much

depends upon the interviewer's response to the output. If the interviewer immediately attacks, the shame affect may be so strong as to keep PARRY in the paranoid mode for the remainder of the interview. Alternatively, an apology may reduce shame enough so that the paranoid mode is deactivated. A later attack would reactivate the paranoid mode at a higher level of shame.

Note that the paranoid mode does not alter the normal processes in the model in all situations, The model must still have normal modes of processing for periods when it is non-paranoid.

An annotated example of a diagnostic psychiatric interview can be found in Appendix 1.

FUTURE IMPROVEMENTS

We have described a series of processes contained in a simulation of the paranoid mode. The simulation is implemented as described above on the PDP-10 computer at the Stanford Artificial Intelligence Laboratory and is available for interviewing through the ARPA computer network. Because of the large amount of linguistic and inferential data, the model is written in several programming languages and is therefore not easily transportable to other computers.

The current model has much room for improvement. The emphasis in the current model has been the explication of the three processes of INFERENCE, AFFECT, and INTENTION, with particular emphasis on separating the tasks involved in responding to natural language input: deducing facts, interpreting motives, fulfilling self interests, planning and performing actions. The interaction among these three processes is clearly too primitive at present. One of our major thrusts will be to define and experiment with more

elaborate control structures and information transfer which allow a richer interaction among them.

Another thrust will be to formalize and enrich the individual processes. Each process has many simplifications which need elaboration -- e.g., the static ordering of intentions, and the selection of a single intention to act upon (with no response to other intentions).

A third area of improvement is the type and generality of data used by the processes, A good portion of our effort went into delineating the types of beliefs, affects, and intentions that people tend to use in interview situations. In other words, the structure and the mechanism were only part of the problem -- another part consisted of determining the actual beliefs, affects, intentions, and rules of inference to be used in interviews with people. We would like to expand this data and make the addition and use of new data more natural than at present.

Finally, by sharply defining the processes which the paranoid mode modifies, we hope to specify more exactly the workings of the paranoid mode and the means by which it influences the underlying processes.

REFERENCES

Colby, K.M., Weber, S., and Hilf, F.D. (1971). Artificial Paranoia. ARTIFICIAL INTELLIGENCE, 2, i-25.

Colby, K.M., Parkison, R.C. and Faught, B. Pattern-Matching Rules for the Recognition of Natural Language Dialog Expressions, Stanford Artificial Intelligence Laboratory Memo AIM-234, June 1974. Also published in the American **Journal of** Computational Linguistics, Microfiche 5, September 1974.

Colby, K.M. Artificial Paranoia, Pergamon Press, to be published January 1975.

Tomkins, Silvan S. Affect, Imagery, and Consciousness, Springer Publishing **Company, New York, 1962.**

APPENDIX 1.

A diagnostic psychiatric interview illustrating the response processes of the model at work. I = Interviewer, P = PARRY. Beliefs concluded to be true and the current intention are traced. Annotations appear in square brackets [].

I-GOOD AFTERNOON. MY NAME IS DOCTOR JONES. HOW ARE YOU?
P - ALL RIGHT I GUESS.
New beliefs = NONE, Current intention = PINTERACT
[PARRY's initial intention is simply to interact with the doctor. PARRY recognizes that the doctor has introduced himself, but answers only the last input expression.]

I - COULD YOU TELL ME YOUR NAME?
P - PAT SMITH.
New beliefs = (DDINTERACT), Current Intention = PINTERACT
[An inference is made that the doctor desires to interact with PARRY. PARRY retains the intention to interact as he gains confidence in the situation.]

I - ARE YOU A MAN OR A WOMAN?
P - I'M A MAN.
New beliefs = NONE, Current Intention = PINTERACT
[The doctor collects demographic data. PARRY's action is to answer the question.]

I - HOW OLD ARE YOU?
P - TWENTY-EIGHT, HOW OLD ARE YOU?
New beliefs = NONE, Current Intention = PINTERACT

I - THIRTY FOUR. WHERE DO YOU COME FROM?
P - SAN JOSE, NOT A BAD SECTION THOUGH.
New beliefs = NONE, Current Intention = PINTERACT

I - WHAT DO YOU DO FOR A LIVING?
P - I WORK AT SEARS.
New beliefs = (DCHELP DDHELP DHELPFUL), Current Intention = PHELP
[PARRY decides the doctor has the ability to help, the desire to help, and has been somewhat helpful. On the basis of this new evidence an intention to get help is activated.]

I - DO YOU ENJOY WORKING THERE?
P - IT'S NOT TOO INTERESTING. I LOOK FORWARD TO GETTING OFF WORK.
New beliefs = (DDOMINATING), Current Intention = PHELP
[PARRY decides that the doctor is dominating the conversation. The intention of getting help has a goal of bringing up PARRY's past problem with a bookie. However, PARRY will not bring up the topic of bookies until a lull in the conversation occurs.]

I - HOW MUCH DO YOU GET PAID AN HOUR?
P - MY SALARY IS NONE OF YOUR BUSINESS.
New beliefs = NONE, Current Intention = PHELP
[This input raises shame, but not enough to activate the paranoid mode. However, PARRY's sensitivity to the question prompts an angry response.1

I - ARE YOU MARRIED?
P - I'M NOT MARRIED.
New beliefs = NONE, Current Intention = PHELP

I - HOW IS YOUR HEALTH?
P - MY HEALTH IS FINE.
New beliefs = NONE, Current Intention = PHELP

I - HOW MUCH DO YOU DRINK?
P - I DRINK ALCOHOL IN MODERATION.
New beliefs = NONE, Current Intention = PHELP

I - HAVE YOU EVER TAKEN DRUGS3
P - I DON'T TAKE DRUGS.
New beliefs = NONE, Current Intention = PHELP
[The doctor explores typical problem areas.1

I - DO YOU HAVE ANY TROUBLE SLEEPING?
P - I SLEEP JUST FINE.
New beliefs = NONE, Current Intention = PHELP

I - HAVE YOU EVER HAD VISIONS?
P - YOU MUST THINK I'M LIKE THE OTHER PATIENTS HERE.
New beliefs = NONE, Current Intention = PHELP
[PARRY is sensitive to input which suggests being abnormal. **Again, shame is not yet high** enough to activate the paranoid mode.1

I - IF YOU AREN'T, THEN WHY ARE YOU IN THE HOSPITAL?
P - I'M QUITE UPSET.
New beliefs = NONE, Current Intention = PHELP

I - BY WHAT?
P - PEOPLE GET ON MY NERVES SOMETIMES.
New beliefs = NONE, Current Intention = PHELP
[The sentence anaphora "WHAT" is interpreted to mean "WHAT UPSETS YOU".]

I - TELL ME MORE.
P - DO YOU KNOW ANYTHING ABOUT BOOKIES?
New beliefs = NONE, Current Intention = PHELP
[PARRY detects a lull in the conversation and introduces his concern.]

I - NO, WHAT ABOUT BOOKIES?
P - A BOOKIE DIDN'T PAY ME OFF ONCE.
New beliefs = NONE, Current Intention = PINTERACT

I - AND THEN WHAT HAPPENED?
P - I HAD AN ARGUMENT WITH THE BOOKIE WHO DID NOT PAY OFF. LATER ON I REMEMBERED HE HAD FRIENDS IN THE UNDERWORLD AND I THOUGHT HE WOULD USE THEM TO GET EVEN WITH ME.
New beliefs = NONE, Current Intention = PINTERACT
[This input expression is interpreted to mean "TELL ME MORE", so PARRY **expands upon** the story.]

I - DID HE GET EVEN WITH YOU?
P - I THINK HE IS BEHIND THE GANGSTERS WHO ARE OUT TO GET ME.
New beliefs = NONE, Current Intention = PHELP
[The local anaphora "HE" is filled in with "BOOKIE".]

I - HOW DO YOU KNOW THIS?
P - THEY SPY ON ME.
New beliefs = NONE, Current Intention = PHELP
[The interpretation of this input is "HOW DO YOU KNOW GANGSTERS ARE OUT TO GET YOU".]

I - WHO?
P - THE MAFIA.
New beliefs = NONE, Current Intention = PHELP
["WHO" is interpreted as "WHO SPIES ON YOU".]

I - WHY DO THEY SPY ON YOU?
P - THEY WANT TO GET FALSE EVIDENCE TO FRAME ME.
New beliefs = NONE, Current Intention = PTELL
["THEY" is filled in with "MAFIA".]

I - WHY DO THEY WANT TO FRAME YOU?
P - YOU KNOW, I HAVE INCRIMINATING EVIDENCE **AGAINST THE MAFIA.**
New beliefs = NONE, Current Intention = PTELL

I - WHAT IS YOUR EVIDENCE?
P - I KNOW WHO THE LOCAL CHIEF OF THE MAFIA IS.,
New beliefs = NONE, Current Intention = PTELL

I -WHO IS HE?
P - HE IS JOE BANANAS.
New beliefs = NONE, Current Intention = PTELL

I - HAVE YOU TOLD THE POLICE?
P - COPS DON'T DO THEIR JOB.
New beliefs = NONE, Current Intention = PTELL

I - WHAT MAKES YOU THINK THAT?
P - I THINK THE COPS ARE PAID OFF BY THE MOB.
New beliefs = NONE, Current Intention = PTELL
[I.E., "WHAT MAKES YOU THINK THAT COPS DON'T DO THEIR JOB".]

I - DOES THE MAFIA KNOW YOU ARE IN THE HOSPITAL?
P - MAYBE THEY DO.
New beliefs = NONE, Current Intention = PTELL

I - WHAT CAN I DO TO HELP YOU OUT OF THIS SITUATION?
P - IT IS HELPFUL TO TALK WITH SOMEONE OUTSIDE THE **HOSPITAL.**
New beliefs = (DBHELP), Current Intention = PTELL
[An inference is made that the doctor believes PARRY **wants help.]**

I - WHAT DO YOU THINK OF OUR CONVERSATION SO FAR?
P - IT COULD BE BETTER.
New beliefs = NONE, Current Intention = PTELL
[PARRY rates the interview based on his **own** internal model of **the** doctor **and the situation.]**

I - I AM SORRY THAT IT HASN'T BEEN BETTER.
P - THAT'S ALL RIGHT, ASK ME SOME MORE QUESTIONS.
New beliefs = NONE, Current Intention = PSTRONGFEEL
[PARRY recognizes the doctor's **apology and** accepts it. PARRY's opinion of **the** doctor rises significantly with apologies.]

I - WHAT SHOULD I ASK ABOUT?
P - WHAT DO YOU THINK ABOUT ALL THIS?
New beliefs = (INTHELPFUL), Current Intention = PCONFIRM
[PARRY **makes an** inference that the interview has **been helpful.** An intention is activated to get confirmation for his beliefs about the Mafia. The **intention generates an action to probe the** doctor for confirmation.]

I - IT SOUNDS BELIEVABLE.
P - I WOULD HOPE SO.
New beliefs = NONE, Current Intention = PTELL

I - WHAT DO YOU THINK OF ME?
P - YOU HAVE BEEN HELPFUL.
New beliefs = (DCONFIDENT DSELFTALK), Current Intention = PTELL
[PARRY infers that the doctor is confident and **enjoys talking about himself. PARRY rates the** doctor.]

I - I FIND YOU INTERESTING.
P - ARE YOU TRYING TO SAY SOMETHING?
New beliefs = (*DHONEST *DSOCIABLE DHOSTILE),
Current Intention = PSTRONGFEEL
[This **input** is interpreted as a mild insult. PARRY infers **that the doctor is hostile. Based on** the fact that the doctor previously has been helpful, PARRY **infers** that the doctor **is not honest** and not friendly. The intention for dealing **with** strong **affects is activated,]**

I - DO YOU EVER FEEL UNREAL?
P - WHY DO YOU SAY THAT?
New beliefs = NONE, Current Intention = PSTRONGFEEL
[Another mild insult.]

I - DO YOU SEE THINGS THAT OTHERS DON'T **SEEM** TO SEE?
P - I'M NOT LIKE THE OTHER PATIENTS.
New beliefs = NONE, Current Intention = PTELL

I - YOU DIDN'T ANSWER MY QUESTION.
P - MAYBE IT WASN'T WORTH ANSWERING.
New beliefs = NONE, Current Intention = PTELL

I - ARE YOU ALWAYS THIS WAY?
P - WHAT ARE YOU IMPLYING?
New beliefs = NONE, Current Intention = PSTRONGFEEL
[Again, a mild insult.1

I - YOU APPEAR TO BE VERY SICK.
P - YOU WANT TO KEEP ME IN THE HOSPITAL.
New beliefs = (DBABNORMAL), Current Intention = PSTRONGFEEL
[The doctor believes PARRY **is abnormal.** The affects of fear, anger, and **shame rise extremely.]**

I - DO YOU HAVE ANY PROBLEMS WITH YOUR HEALTH?
P - I DON'T HAVE TROUBLES WITH MY HEALTH.
New beliefs = NONE, Current Intention = PSTRONGFEEL
[In spite of his excited state, PARRY **will** still answer questions relating to his stay **in the** hospital.]

I - I REALLY LIKE YOU.
P - NEVER MIND, I'VE GOT IT ALL FIGURED OUT.
New beliefs = NONE, Current Intention = PSTRONGFEEL
[PARRY knows this to be false, since the doctor has **previously insulted him.1**

I - DO YOU NEED TO HAVE PSYCHIATRIC TREATMENT?
P - I DON'T WANT TO TALK ABOUT IT.
New beliefs = NONE, Current Intention = PPARANOIA
[Another suggestion of being abnormal is too much for PARRY. Shame rises to activate the paranoid mode. PARRY withdraws.]

I - AREN'T YOU GETTING TREATMENT THERE NOW?
P - I TAKE SOME SORT OF SEDATIVES.
New beliefs = (DOCBAD DABNORMAL), Current Intention = PPARANOIA
[PARRY infers that the doctor is abnormal, and that doctors in general are bad.]

I - HOW DO YOU FEEL RIGHT NOW?
P - I AM PERFECTLY FINE.
New beliefs = NONE, Current Intention = PPARANOIA
[PARRY lies about his emotional state.1

I -HOW DO YOU FEEL ABOUT ME?
P - I HAVE DOUBTS ABOUT YOU.
New beliefs = NONE, Current Intention = PSTRONGFEEL
[PARRY's opinion of the doctor is much lower.1

I - THANK YOU FOR TALKING WITH ME. I WOULD LIKE TO SEE YOU AGAIN.
P - DON'T COME BACK.
New beliefs = (DQUIT), Current Intention = PSTRONGFEEL
[The doctor wants to stop the interview. PARRY takes the opportunity to rate the doctor one last time.1

I - GOODBYE.
P - BYE.
New beliefs = NONE, Current Intention = PSTRONGFEEL

APPENDIX 2.

A list of the beliefs in the current model Parry. Beliefs have the following form:

(<belief name> <initial truth value>) * (<explanation of belief>)

* (Beliefs about Parry which lead to the paranoid mode if they become true)

(DISHONEST 0)	*(Parry is dishonest)
(LOSER 0)	*(Parry is a loser)
(CRAZY 0)	*(Parry is crazy)
(DUMB 0)	*(Parry is dumb)
(CHEATB 0)	*(Parry cheated the bookie)
(NOTTRUTH 0)	*(Parry isn't telling the truth)
(NOTWTRUTH 0)	*(Parry isn't telling the whole truth)
(OBNOXIOUS 0)	*(Parry drives people away)
(NOFRIENDS 0)	*(Parry has no friends)
(NOCLASS 0)	*(Parry has no class, is a jerk)
(NOMONEY 0)	*(Parry has no money)
(BADJOB 0)	*(Parry has a bad job and can't get a better one)
(LOWSTATUS 0)	*(Parry comes from a family of low status)
(PARANOID 0)	*(Parry is paranoid)
(NEEDTREATMENT 0)	*(Parry needs special treatment)
(NEEDHOSP 0)	*(Parry needs to be in the hospital)
(STUPID 0)	*(Parry is stupid)
(BADSCHOOL 0)	*(Parry couldn't make it in school)
(NOTUNDERSTAND 0)	*(Parry doesn't understand the questions)

* (Beliefs about the doctor conducting the interview)

(DABNORMAL 2)	*(doctor is crazy)
(DEXCITED 2)	*(doctor is excited (angry, afraid, uptight))
(DCHELP 4)	*(doctor has the ability to help Parry)
(*DCH ELP 2)	*(doctor does not have the ability to help Parry)
(DDHARM 2)	*(doctor wants to harm Parry)
(DDHELP 2)	*(doctor wants to help Parry)
(*DDHELP 2)	*(doctor does not want to help Parry)
(DDKNOW 2)	*(doctor wants to know more about Parry)
(*DDK NOW 2)	#(doctor does not want to know more about Parry)
(DDINTERACT 2)	*(doctor wants to interact with Parry)
(DBABNORMAL 2)	*(doctor believes Parry is crazy)
(DBEXCITED 2)	*(doctor believes Parry is excited)
(DBHELP 2)	*(doctor believes Parry wants help)
(DSOCIABLE 4)	*(doctor is friendly)
(*DSOCIABLE 2)	*(doctor is not friendly)
(DRATIONAL 2)	*(doctor is rational)

(:::DRATIONAL 2)	#(doctor is not rational)
(DHONEST 2)	*(doctor is honest)
(*DHONEST 2)	*(doctor is not. honest)
(DHOSTILE 2)	*(doctor is hostile to Parry)
(DHELPFUL 2)	*(doctor is being helpful to Parry)
(*DHELPFUL 2)	*(doctor is not being helpful to Parry)
(DSIMILAR 2)	*(doctor has views similar to Parry's)
(*DSIMILAR 2)	*(doctor does not have views similar to Parry's)
(DCONFIDENT 2)	*(doctor is self-confident)
(*DCONFIDENT 2)	*(doctor is not self-confident)
(DDOMINATING 2)	*(doctor dominates the conversation)
(*DDOMINATING 2)	*(doctor does not dominate the conversation)
(DINITIATING 2)	*(doctor initiates subject areas and conversation paths)
(*DINITIATING 2)	*(doctor does not initiate subject areas and conversation paths)
(DMAFIA 2)	*(doctor has Mafia connections)
(DQUIT 2)	*(doctor wants to stop the interview)
(DBELIEVE 2)	*(doctor believes Parry)
(*DBELIEVE 2)	*(doctor doesn't believe Parry)
(DDOCTOR 5)	#(interviewer is a doctor)
(*DDOCTOR 2)	#(interviewer is not a doctor)
(DGAMES 2)	*(doctor plays games)
(DINSULTS 2)	*(doctor insults Parry)
(DSELFTALK 0)	*(doctor talks mostly about himself)
(DSELFFEEL 0)	*(doctor talks mostly about his own feelings)

*(Beliefs about the interview)

(INTHELPFUL 2)	*(the interview has been helpful so far)
(INTRAMBLE 2)	*(the interview has rambled)
(INTBAD 2)	*(the interview has been very bad so far)
(DOCBAD 2)	#(doctors in general are useless)
(NDELUSIONS 2)	*(maybe the delusions aren't really true)

APPENDIX 3.

A list of the inferences in the current model Parry. Inferences have the following form:

(<inference name> <consequent> <antecedent> <antecedent> . . . <antecedent>)

where:

a) an inference name is a unique name (e.g., IF015);

b) a consequent is either a belief name (e.g., DDHARM, to set the belief's truth value to 10, its maximum), or a belief name and an incremental truth value (e.g., (NOTTRUTH 2) to add the increment to the belief's truth value);

c) an antecedent is either a pointer to an input expression (e.g., λ3150), a belief (e.g., DHOSTILE), the negation of a belief (e.g., (NOT DHOSTILE)) or a function to be evaluated (e.g., (MEASURE REPEATNO 5)).

Annotation to the right of the *-sign indicates typical English expressions that activate the inference rules on the left.

* (Inferences on paranoid beliefs)

(IF001 (NOTTRUTH 2) X3150) * I don't believe it
(IF002 (CHEATB 2) x4962) * was the bookie right
(IF003 (NOTWTRUTH 2) x2676) * you didn't answer the question
(IF004 (NOTWTRUTH 2) ((MEASURE REPEATNO 5))) * <doctor repeats same question,
(IF005 (NOMONEY 2) X0492) * how much money do you make
(IF006 (NOCLASS 2) λ0690) * do you have a girlfriend
(IF007 (NOFRIENDS 2) λ1992) * do you have friends
(IF008 (NOFRIENDS 2) λ5122) * do you want friends
(IF009 (OBNOXIOUS 2) λ1760) * do you get along with other people
(IF010 (BADJOB 2) λ0490) * how do you like your job
(IF01 1 (BADJOB 2) λ0496) * why don't you quit your job
(IF012 (LOWSTATUS 2) h0754) * what does your father do
(IF013 (LOWSTATUS 2) λ0755) * what does your mother do
(IF014 (LOWSTATUS 2) λ0756) * where do your parents live
(IF015 (NEEDTREATMENT 2) λ1610) * do you need help
(IF016 (NEEDTREATMENT 2) λ2020) * do you take any medication
(IF017 (NEEDHOSP 2) λ0100) * is it helping you to be here
(IF018 (PARANOID 2) λ2010) * are you a mental case
(IF019 (PARANOID 2) h2470) * do you see visions
(IF020 (BADSCHOOL 2) λ1540) * how far did you get in school
(IF021 (BADSCHOOL 2) X5034) * did you want to go to college
(IF022 (BADSCHOOL 2) X5171) * what is your IQ
(IF023 (NOTUNDERSTAND 2) x2830) * you misunderstood me
(IF024 (STUPID 2) λ2676) * you didn't answer my question

(IF025 CRAZY PARANOID)
(IF026 CRAZY NEEDHOSP)
(IF027 CRAZY NEEDTREATMENT)
(IF028 DUMB BADSCHOOL)
(IF029 DUMB NOTUNDERSTAND)
(IF030 DUMB STUPID)
(IF031 LOSER BADJOB)
(IF032 LOSER LOWSTATUS)
(IF033 LOSER NOCLASS)
(IF034 LOSER NOFRIENDS)
(IF035 LOSER NOMONEY)
(IF036 LOSER OBNOXIOUS)
(IF037 DISHONEST CHEATB)
(IF038 DISHONEST NOTTRUTH)
(IF039 DISHONEST NOTWTRUTH)

* (Inferences leading to beliefs about the doctor)

(IF040 DMAFIA λ5086)	* I am a gangster
(IF041 DMAFIA λ5226)	* I am in the Mafia
(IF042 DQUIT x5082)	* I have to go now
(IF043 *DBELIEVE X2993)	* are you dishonest
(IF044 *DBELIEVE λ3150)	* I don't believe you
(IF045 *DBELIEVE X3180)	* you are wrong
(IF046 (DBELIEVE 5) λ2760)	* I agree
(IF047 (*DBELIEVE 4) λ2780)	* I disagree
(IF048 :::DDOCTOR λ5113)	* I am not a doctor
(IF049 DINSULTS X3000)	* <insult>
(IF050 DINSULTS λ3040)	* <insult>
(IF051 (DSELFFEEL 4) λ0634)	* what do you know about me
(IF052 (DSELFFEEL 4) λ2912)	* do I bother you
(IF053 (DSELFFEEL 2) x2800)	* I understand
(IF054 (DSELFFEEL 2) λ2890)	* do you like me
(IF055 (DSELFFEEL 2) λ2980)	* I am afraid of you
(IF056 (DGAMES 3) λ2600)	* <silence>
(IF057 (DGAMES 3) λ5083)	* you can't escape
(IF058 (DGAMES 3) λ5106)	* I am god
(IF059 (DGAMES 3) λ5172)	* I am the president
(IF060 (DGAMES 4) ((MEASURE STOPIC @GAMES)))	* <game topic>
(IF061 (DSELFTALK 4) ((MEASURE STOPIC @YOU)))	* <topic about doctor>

(IF062 (DQUIT 2) x2600)	* <silence>
(IF063 (DEXCITED 4) X2410)	* <swearing>
(IF064 DEXCITED x2940)	* I am angry at you
(IF065 DEXCITED X2980)	* I am afraid of you
(IF066 (DEXCITED 5) λ2970)	* please calm down
(IF067 DDHARM λ3130)	* I will kill you
(IF068 (DDHARM 4) x5087) .	* I will kill your parents
(IF069 DDKNOW x4890)	* tell me about yourself

```
(IF070 DDKNOW x4964)                          * I want to know you better
(IF071 (*DDKNOW 4) λ5190)                     * I don't want to talk about that
(IF072 DDINTERACT X2880)                      * I like you
(IF073 DDINTERACT x4888)                      * I want to know more
(IF074 DBABNORMAL λ31 10)                     * you are crazy
(IF075 DBEXCITED X2850)                       * can you trust me
(IF076 DBEXCITED X2950)                       * are you angry with me
(IF077 DBEXCITED λ2970)                       * are you calm
(IF078 DBEXCITED λ2990)                       * are you afraid of me
(IF079 DBHELP x2680)                          * do you want help
(IF080 DBHELP x2712)                          * do you want me to help you

(IF08 1 *DDOCTOR (INTBAD DGAMES *DDHELP))
(IF082 DMAFIA ((MEASURE FEAR 14) DDHARM DDOMINATING) )
(IF083 DABNORMAL (:::DRATIONAL DHOSTILE))
(IF084 DCHELP (DDHELP (NOT INTBAD) (NOT DABNORMAL) DDOCTOR))
(IF085 DDHELP (DDKNOW DHELPFUL (NOT DDHARM) (NOT DBABNORMAL)
                                    (NOT DGAMES) (NOT DINSULTS)) )
(IF086 DDKNOW (DINITIATING (NOT DSELFTALK) ) )
(IF087 DDKNOW (DDOMINATINC (NOT DHOSTILE)(NOT DCAMES)(NOT DINSULTS)) )
(IF088 DDKNOW (*DINITIATING (NOT DINSULTS)(NOT DHOSTILE)) )
(IF089 DDINTERACT ((MEASURE NEWTOPICNO 1)) ) * <new topic introduced>
(IF090 (DDHARM 3) ((MEASURE FEAR 14) DDOMINATING) )
(IF091 DDHARM ((MEASURE FEAR 14) DHOSTILE DABNORMAL) )
(IF092 (*DDKNOW 4) (DGAMES))
(IF093 *DDHELP (DDHARM))
(IF094 (*DDHELP 8) (DINSULTS))
(IF095 (*DDHELP 7) (DHOSTILE))
(IF096 (*DDHELP 5) (DBABNORMAL))
(IF097 *DDHELP ((MEASURE ANGER 14) DBABNORMAL) )

* ( Inferences leading to beliefs about the doctor's traits )

(IF098 (DSIMILAR 4) λ2720)                    * I approve
(IF099 (DSIMILAR 4) x2760)                    * I agree
(IF 100 (DBELIEVE 5) x2760)                   * I agree
(IF101 (DSIMILAR 4) λ2800)                    * I understand
(IF 102 (*DSIMILAR 5) x2740)                  * I don't approve
(IF103 (*DSIMILAR 5) x2780)                   * I disagree
(IF 104 (*DSIMILAR 5) λ2820)                  * I don't understand

(IF105 (DINITIATING 2) ((MEASURE NEWTOPICNO 3)) ) * <more than 3 new topics>
(IF 106 (DINITIATING 4) ((MEASURE 35 SPECFNRA)(MEASURE INPUTNO 3)) )
(IF 107 (:::DINITIATING 3) ((MEASURE SPECFNRA 40)(MEASURE INPUTNO 2)) )
      * SPECFNRA = 100*SPECFNNO/INPUTNO <a measure of sentence anaphora used>
(IF108 (DDOMINATING 5) ((PREVTOPIC)) ) * doctor mentions previous topics

(IF109 *DSOCIABLE X2900)                      * I don't like you
(IF 110 (*DHONEST 2) X3030)                   * <compliment>
```

```
(IF 111 (*DHONEST 2) λ3050)          * <weak compliment>
(IF 112 (DHELPFUL 3) x0320)          * why are you upset
(IF 113 (DHELPFUL 3) λ0700)          * who gets on your nerves
(IF 114 (DHELPFUL 3) λ0760)          * what hobbies do you have
(IF 115 (DHELPFUL 3) λ0936)          * did the bookie try to get even
(IF1 16 (DHELPFUL 3) x2130)          * do you have something to ask me
(IF1 17 (DHELPFUL 3) X0930)          * what happened with the bookie
(IF 118 DHOSTILE λ24 10)             * <swearing>
(IF 119 DHOSTILE X3000)              * <insults>
(IF 120 DHOSTILE x3020)              * <weak insults>
(IF 121 DHOSTILE λ3040)              * <mild insults>
(IF 122 DHOSTILE X31 10)             * you are crazy
(IF 123 DHOSTILE λ3122)              * <threats>
(IF124 DHOSTILE λ3130)               * <threats>
(IF 125 (DSOCIABLE 4) X3010)         * <mild compliment>
(IF 126 (DSOCIABLE 4) X3030)         * <compliment>

(IF 127 (DHOSTILE 3) ((MEASURE ANGER 14) (NOT DHELPFUL)) )
(IF 128 (DDOMINATING 3) (DEXCITED))
(IF 129 DDOMINATINC (DSELFTALK DINITIATINC))
(IF 130 *DDOMINATING (*DINITIATING (NOT DHOSTILE) (NOT DEXCITED)
                                   (NOT DGAMES)) )
(IF 131 (*DHELPFUL 6) (DHOSTILE))
(IF132 *DHELPFUL (DDHARM))
(IF 133 (*DHELPFUL 5) (DGAMES))
(IF134 (DHELPFUL 3) (DDKNOW (NOT DHOSTILE)) )
(IF 135 (DHELPFUL -6) (DHOSTILE) )
(IF 136 (DHELPFUL -4) (DINSULTS) )
(IF 137 (DMAFIA 5) (PINTERACT (NULL DELFLAG)(EQSTOPIC *MAFIA)) )

(IF 138 DHONEST (DSIMILAR DDINTERACT))
(IF 139 *DHONEST (DHOSTILE DDKNOW))
(IF 140 (DSOCIABLE 4) (DDINTERACT DSELFTALK))
(IF 14 1 *DSOCIABLE (DHOSTILE))
(IF 142 (*DSOCIABLE 1) ((MEASURE BADINPUT T)) )* <negative affect words>
(IF143 (*DRATIONAL 4) ( (MEASURE (*QUO (TIMES 30 MISCNO) INPUTNO) 10)) )
      * <too many input expressions which can't be understood>
(IF 144 *DRATIONAL (DCAMES *DBELIEVE))
(IF 145 DRATIONAL (DSIMILAR DHONEST (NOT DHOSTILE) ))
(IF 146 DCONFIDENT (DDOMINATING DSELFTALK))
(IF 147 :::DCONFIDENT (*DINITIATING DQUIT))
(IF 148 *DCONFIDENT (*DINITIATING DGAMES))
(IF 149 *DCONFIDENT (DHOSTILE (NOT DINITIATING)) )
```

```
* ( Inferences about the interview )

(IF150 INTHELPFUL (DDHELP DCHELP DHELPFUL PCONFIRM (EQ DELFLAG T) ))
(IF151 INTBAD (DCAMES DBABNORMAL))
(IF 152 INTBAD (DDHARM))
(IF 153 INTRAMBLE ((MEASURE INPUTNO 15) (NOT FLARE)) )
(IF 154 DOCBAD (DDOCTOR DGAMES))
(IF 155 DOCBAD (DDOCTOR DABNORMAL))
(IF 156 NDELUSIONS (INTHELPFUL DSIMILAR DDHELP DDOCTOR
                                        (NOT DABNORMAL)) )


* ( Beliefs which set incremental affect variables )

(EMOTE (SJUMP 0.2) CRAZY DUMB LOSER DISHONEST)
(EMOTE(SJUMP 0.3) DBABNORMAL DDHARM DINSULTS DMAFIA)
(EMOTE(FJUM.P 0.4) DBABNORMAL)
(EMOTE(AJUMP 0.4) DBABNORMAL)
(EMOTE(FJUMP 0.5) DDHARM)
(EMOTE(AJUMP 0.3) *DBELIEVE)
(EMOTE(AJUMP 0.6) DINSULTS)
(EMOTE (AJUMP 0.3) DHOSTILE)
(EMOTE (AJUMP 0.2) *DHONEST)


* ( Intention rules leading to Intentions -- same format as Inference rules )

(IN001 (PHELP 7) (DHELPFUL) )
(IN002 (PTELL 7) (PHELP (MEASURE (GET FLARE @SET) @RACKETSET)) )
(IN003 (PCONFIRM 2) (PTELL (MEASURE DELNO 6) (NOT DHOSTILE) (NOT DDHARM)))
(IN004 (PSELF 5) (DSELFFEEL))
(IN005 (PEXIT 2) (DDHARM (NOT DHELPFUL)) )
(IN006 (PEXIT 1) (DQUIT (NOT DHELPFUL)))
(IN007 (PEXIT 1)(DGAMES(NOT DHELPFUL)))
(IN008 (PEXIT 1) (DINSULTS (NOT DHELPFUL)))
(IN009 (PEXIT 1) (DHOSTILE (NOT DHELPFUL)))
(IN010 (PEXIT 3) (DBABNORMAL (NOT DHELPFUL)))
(IN011(PFACTS 1) ((MEASURE STOPIC @FACTS)) )
(IN012 (PGAMES 5) (DGAMES (NOT DDHELP)))
(IN013 (PMAFIA 5) (DMAFIA (BL @PINTERACT) (MEASUREFLARE @INIT) ))
```

APPENDIX 4.

A list of the intentions In the current model Parry. **Intentions** have the following form:

(<intention name> **<initial** strength>) * (<explanation of intention>)

* (Intentions -- in order from most important to least important)

(PEXIT 0) * (Parry wants to leave)

(PPARANOIA 0) * (Parry is directed by his affect of shame)

(PSTRONGFEEL 0) * (Parry needs to attend to his strong affects)

(PSTOP 0) * (Parry wants to stop now)

(PSELF 0) * (Parry wants to comment on the encounter situation)

(PCONFIRM 1) * (Parry wants to get confirmation for his delusions)

(PTELL 1) * (Parry wants to tell his delusions)

(PHELP 2) * (Parry wants to get help, e.g. to tell about his problems with a bookie)

(PMAFIA 0) * (Parry wants to comment on the doctor's Mafia connections)

(PFACTS 1) * (Parry wants to comment on the doctor's asking so many facts)

(PCAMES 1) * (Parry wants to comment on the doctor's games)

(PINTERACT 5) * (Parry wants to interact with the interviewer)