STANFORD ARTIFICIAL INTELLIGENCE LABORATORY
MEMO AIM-244

STAN-CS- 74 - 457

# TEN CRITICISMS OF PARRY

BY

KENNETH MARK COLBY

SUPPORTED BY

NATIONAL INSTITUTE OF MENTAL HEALTH

SEPT 1974

COMPUTER SCIENCE DEPARTMENT
School of Humanities and Sciences
STANFORD UNIVERSITY

STANFORD ARTIFICIAL INTELLIGENCE LABORATORY
MEMO-AIM244

SEPTEMBER 1974

COMPUTER SCIENCE OEPARTMENT
REPORT NO. STAN-CS-74-457

# TEN CRITICISMS OF PARRY

Kenneth Mark Colby (1)

Abstract:

Some major criticisms of a computer simulation of paranoid processes (PARRY) . are revieued and di scussed.

(1) Senior Research Associate, Department of Computer Science, Stanford University

This research is eupported by Grant PHS MH06645-13 from the National Institute of Mental Health and in part, by Research Scientist Award (No. 1-K05-K-14,333) from the National Institute of Mental Health,

# TEN CRITICISMS OF PARRY

Kenneth Mark Colby

Much of the Artificial Intelligence community is now aware of a computer simulation of paranoid processes developed by the Colby group at Stanford. The model (called PARRY) has been available for interviewing on the ARPA network and thousands of interviews have been conducted 'with several versions of the model. During the long period of development of the model, we have been auare of the limitations of various alternative programming approaches to designing an algorithm capable of conducting useful non-trivial dialogue in natural language. Colleagues, associates, and students have volunteered a number of criticisms of the model. Since criticisms can be endless, I shall restrict the discussion to only those which we consider serious, reasoned, and well-founded.

Workers in A-I come from different intellectual traditions. One's intellectual background influences one's image of what a model is and how it should function. Those from mathematical and logical backgrounds like to see lots of deductive inference; those from physics and chemistry like to see laws; those from the life sciences like to see complexity, growth, and development represented in a model. It is important that we recognize and respect the traditions and philosophies of both the demonstrative and empirical sciences. Those raised on a Euclidean model of knouledge seek to understand' phenomena using a few definitions and axioms, a feu rules of inference, long chains of inference, and deductive consistency. Some aspects of experience yield to this approach but many, especially in the case of living organisms, do not.

Everyone realizes that a model represents a simplification and an idealization. In constructing a model, only a few variables are selected as centrally relevant while the rest are neglected as secondary or unknown. Only a few relations between the relevant variables are introduced. Thus a model does not match exactly that which it models in all details. It is partial in that only some aspects of the referent system are represented and it is an approximation in that it is limited in depth and not free of error. A model is an idealization in that it may utilize abstractions and it may possess perfect properties knoun to be lacking in its natural counterpart. Hence the model's knowledge is not as extensive as that of a person and it possesses a perfect memory unencumbered by inhibitory processes. We can allow ourselves this idealization of perfect memory because ue are not studying, for example, memory decay, since we do not consider it to be a pertinent variable in paranoid processes.

These points are discussed in greater detail in a forthcoming monograph (Colby, 1974).

I shall list ten major criticisms of the model which have come to our attention and attempt replies to each.

CRITICISM #1:

PARRY is simply a stimulus-response model. It recognizes something in the- input and then just responds to it without "thinking" or inferring. The model should interpret what it sees and engage in more computation than execution of a simple reurite or production rule.

REPLY:

It is true that in early versions of the model many of the responses consisted of simple rewrites, e.g. when the input consisted of "Hello", the output response was "Hi" and no rules other than of the type "see x, say y" were involved. (Colby, Weber, and Hilf, 1971). But as we began to improve and extend the model, this type of response disappeared. PARRY no longer consists

of a single program: rather it is a system of programs. In the current version, the model consists of two modules, one for recognition of natural language expressions. and one for response. Once the recognizer decides what is being said, the response module, using a number of tests and rules, decides how to respond. The output action of the model is now a function of input, beliefs, affects, and intentions. Thus a 'Hello' no longer receives an automatic "Hi' but may receive a variety of responses depending on a large number of conditions, including a "model" of the interviewer which PARRY builds up during the interview. This representation of the interviewer involves making inferences about his competence, his helpfulness, etc.

CRITICISM #2:

PARRY'S language recognition processes do not analyze natural language input sufficiently. They only try to match patterns and thus they are naive and simplistic linguistically.

REPLY:

PARRY does not utilize a grammar in processing its input of everyday conversational English. Whereas grammar-based parsers may be sophisticated linguistically, they are too fragile to operate satisfactorily in real-time interviews allowing unrestricted English. PARRY'S language-recognition module uses pattern-matching rules which attempt to characterize input expressions by progressively transforming them into patterns which match, completely or fuzzily, abstract stored patterns. The power of this approach lies in its ability to ignore recognized and unrecognized words and still grasp the meaning of the message. (Colby, Parkison, and Faught, 1974).

Our problem was not to develop or apply a linguistic theory nor to assert hypotheses about how people process language. Our problem was to design a working algorithm which recognizes what is being said in a dialogue in order to make a linguistic response such that a sample of I-O pairs from the paranoid model is judged similar to a sample of I-O pairs from paranoid patients. Seeking effectiveness in real-time with unrestricted input, ue took a straightforward A-I approach to the problem. This approach has proved to be adequate for our purposes.

CRITICISM #3:

PARRY's performance constitutes an illusion. The model's data-base knowledge is too limited to represent adequately all that a person knows. Because the model can answer a few questions well, people (having many tacit expectations and presuppositions) are easily fooled into believing PARRY, is capable of answering the great variety of questions a person is capable of answering. People will assume there is much more there than there really is. Thus PARRY represents a mirage, a conjurer's trick in which the audience is led to believe something is true when it is not.

REPLY:

One of Descartes' tests for distinguishing man from machines uas that the latter 'did not act from knowledge but only froni the disposition of their organs" . (Descartes' other test concerned linguistic variety). Granted that a model of a psychological process should contain knowledge, the questions become, how much knowledge and how is it to be represented?

Since a model is a simplification, it has boundary conditions. A model of a paranoid patient is a model of being paranoid, 'being a patient, and being a person. PARRY does reasonably well in the first two of these "beings". It fails in the third because of limited knowledge. How can we decide what the model should know? It is theoretically trivial to add tomes of facts to the data base, but this seems to be what some A-I critics uant, The fact that PARRY can discuss some topics rather well indicates it is doing the right things in these domains and could do well in other domains that are functionally similar. Simply adding facts without improving the algorithm can lead to a degradation of performance as experience with belief-system simulations and theorem-proving programs has shown.

More important than sheer number of facts is how they are organized, how they are represented, and how they are handled by the processing rules to contribute to the characteristic performance of the model. Some seem happy to know there are fixed propositions or "frames" in the data-base which can be consulted in answering questions. Even if a model can answer 50 questions about a topic using rewrite rules, some would say the model does not really "know" anything about the topic. The procedural- declarative argument has no end in sight. It seems to be a matter of personal style and efficiency.

PARRY is not a literal copy of a total person. The test of adequacy here is not Turing's machine-question-"which is person and which is machine?" This is not a stringent test, since the criteria for distinguishing what is human behavior over a teletype have not been systematically worked out, i.e., almost anything is accepted as being human. (Colby, Hilf, Weber, and Kraemer, 1972). PARRY is not the real thing: it is a model, a simulation, an imitation, a mind-like artifact, an automaton, synthetic and artificial. The real thing, a living person, is characterized by such great logical complexity, inhomogeneity of class, and individuality that a strategy of simplification is called for.

**CRITICISM #4:**

PARRY models paranoid behavior without modelling the underlying mechanisms of paranoid processes. Because the I-O behavior of PARRY is indistinguishable from the I-O behavior of paranoid patients, it does not mean that the same mechanisms are involved.

**REPLY:**

This is so true as to be an A-I truism. When the inner mechanisms of a system are inaccessible to observation, one must make plausible guesses as to what is going on. These guesses represent analogies. They are not to be taken as the "same" mechanisms. If we kneu the "real" mechanisms, there would be no need to posit analogies about a hidden reality. We try to design structures to fill in more and more of the black box. Further empirical tests and experiments are necesary to increase the plausibility of the analogy proposed. Successful predictions and pragmatic usefulness increase the acceptability of the model to the relevant expert community or communities.

We can never know with certainty whether a model is "true". If it is consistent uith itself and with the data of observation, then it is valuable cognitively and pragmatically. Such coherence is not a definition, of truth but a criterion for truth.

An expert community has various criteria for acceptability of a model. Sometimes it is demanded that a model provide an explanation. What constitutes an explanation may range from describing causes to making intelligible the connections betueen input and output. An extreme view is that science does not explain anything; A is simply interpreted in terms of B and B in terms of C, etc.

A pragmatic criterion for a model is whether it represents a a workable possibility. Can it be tested and measureably improved a's a result of these tests? That is, is there an evaluation procedure for cumulative progress? In the case of PARRY, the answer to these questions is "yes". (Colby and Hilf, 1974).

CRITICISM #5:

PARRY is an ad hoc model. It is designed after the fact to fit a limited set of special cases and lacks generality.

**REPLY:**

Sometimes this criticism is level led at the language-recognition processes and sometimes at the scope of the model. The language recognizer of PARRY is a pattern-matcher. The surface English input expressions are transformed into more abstract patterns which are matched against stored patterns. The many-to-one tranformation involves synonymic-translations and

word-c I asses. Thus the language-recognizer has some generality in that these processes can be used by any "host" system which takes natural language input.

It is true that PARRY is circumscribed. It "explains" the data it was designed to explain. One wants to achieve at least this degree of explanatory power in a model. But can it predict a neu fact or fit a new fact discovered in some other way? This view sees ad hocness, not as a property of a model, but as a relation between two consecutive models or theories. Does PARRY have some novel consequence compared to its predecessor? One trouble is that predecessor formulations explaining paranoia have been so vaguely stated as to be untestable. The theory embodied in the mode I has nove I consequences compared to other formulations.

For examp I e, the theory posits that the paranoid mode of thought involves symbol-processing strategies which attempt to forestall or minimize the affect of humiliation. A novel consequence of this theory is that if a person were desensitized to the negative affect of humiliation, he uould be less prone to uti I ize the strategies of the paranoid mode.

CRITICISM #6:

PARRY'S paranoid behavior is strictly the result of canned paranoid-like responses. Granted that PARRY is diagnosed as paranoid by expert judges, this diagnosis is not a consequence of the theory embodied in the model but is simply produced by the model's canned replies which are linguistically paranoid in nature.

REPLY:

This is a weighty criticism because it implies that the theory of humiliation and the rules of the model are excess baggage. The made-up output repl ies are so typical of paranoid verbal responses that they alone might be sufficient to simulate paranoid interactions.

Given that a model had a list of paranoid-like responses, it would stil I need some mechanism or rules for selecting which response to output in reply to a specific input. Experiments have shown that random selection from this list results in an inadequate performance. For example, on a dimension of "thought disorder" on a 0-9 scale, (0 means zero amount and 9 means a large amount), a random model received a mean rating of 5.94 from expert psychiatrists. Patients rated by the same judges received a mean rating of 2.99 whereas a version of PARRY was rated at 3.78. (Colby and Hilf, 1974).

Little is known about how to generate surface English which is appropriate to the input and phrased in a characteristic style. Segment-by-segment generation or even word-by-word generation would be preferable to outputting canned sentences as long as the rules posited for the paranoid mode were somehow called into play in the generation process. (Fortunately no one has demanded that PARRY generate uords letter-by-letter to account for alternative 'spellings). Since generation of natural language output represents one of the major shortcomings of the model, we are at present attempting to couple the generation more closely with the model's theory.

CRITICISM #7:

The model, even if successful as a simulation, is useless. Does it teach us anything about paranoia? Can it be used to help patients suffering from paranoid disorders?

REPLY:

The model represents an attempt to make intelligible paranoid processes in expl ici t symbol-processing terms. A model of psychopathology in which the mind is in error about some of its own processes has implications for prevention, reduction, and cure of disorder. PARRY intersects two expert communities consisting of researchers in artificial intelligence and

cl inicians in psychiatry. Clinicians are practical men uho are interested in technological applications.

If the disorder is at the "hardware" level of brain pathology, then the application of symbol-processing techniques might be of I ittle use. But i f there is reason to believe the disorder is at the program level of I earned, acquired strategies, then attempts at re-programming through symbolic-semantic techniques are worth considering. At present clinicians have great difficulties treating paranoid disorders. Often the treatment is I imi ted to tranqui lizing drugs. For a cl inician practicing behavior therapy, the model's theory suggests desensitizing the patient to humiliation, a technique which has been successful uith other negative affects such as anxiety. For those practicing psychotherapy, the model's theory suggests exploring the topics of humiliation and self-censure in the hope of helping the patient to reject his judgements of himself as inadequate. Judging whether these treatments are effective would depend on clinical evaluations.

A practical application for PARRY lies in its use as a train'ing aid. Medical students in psychiatry, students in clinical psychology, and psychiatric residents can practice intervieuing PARRY for hours before they "practice" on human patients. They can learn what sorts of input expressions upset the model and lead to uitholding of information or breaking off the doctor-patient relationship.

**CRITICISM #8:**

PARRY does not tell us what is the cause of paranoid thinking, Effective treatment requires we know the cause of a disorder.

**REPLY:**

PARRY does not account for how a system got to be that uay; it describes only hou the system now works. An ontogenetic or morphogenetic model would show how a normal system became that way as a result of its experience over time.

It is not true that to have effective treatments one must know the cause of a di sorder. Illnesses involve loops and circles which, if broken anywhere, can lead to relief of the disorder even uhen the mechanism of action of the treatment is not understood. Common examples of successful treatments for illnesses of unknoun causes are insulin in diabetes, digitalis in congestive heart failure, colchicine in gout, and lithium in mania.

CRITICISM #9:

The tests PARRY has passed are not severe enough. If a model passes a validation test, it might not be because it is a good model but because the test is weak.

**REPLY:**

Our strongest test involves having judges rate interviews with versions of the model and uith paranoid patients. We utilize statistical measures to see how closely the model's performance matches that of the patients and how much better it performs than previous model-versions. A recent study showed that on the dimension of linguistic comprehension independent raters gave PARRY2 a mean rating of 5.48 on a scale of 0-9. (Colby, Hi lf, Wi ttner, Faught, and Parkison, 1974). A previous version of PARRY received a mean rating of 5.25. This improvement is significant at the 0.05% level. But the model is still far from the 7.42 rating received by the patients. The rating groups ( psychiatrists and graduate students ) have been shown to be reliable, i.e., there is agreement both within groups of raters and be tueen groups.

Stronger tests are certainly needed, and we would welcome suggestions along these lines. Are there validation tests others have used which might be suitable for PARRY? In the past most models have relied on face validity. To i mprove a mode l measurcab ly, ue need better tests and statistical measures. One weakness of A-I as a field is that many of its models have not been

sufficiently subjected to empirical tests.

CRITICISM #10:

PARRY is excessively crude, sketchy, and immature as a model. Such theoretical models can be premature for a field and can turn out to be irrelevant or counterproductive. We should collect more data about naturally-occurring paranoia before attempting model construction.

REPLY:

No one really knows when to begin theorizing. Even facts are now believed to be heavily theory-laden, whether their collector realizes it or not. One of the perils of model building is that data used to test a model may demolish it. A model is only sufficient unto the day.

If PARRY is not acceptable, then one accepts some rival formulation ( a current one is "paranoia represents the transformation of love into hate"), or one accepts nothing and waits. Waiting for perfection can be paralyzing to a field, especially one devoted to patients uho need help.

As a simplification, PARRY is perhaps too simple at the moment. In constructing a model, one strives for something simpler than the "real" referent system which is difficult to understand or manipulate. But one wants to retain the important features characteristic of the natural counterpart. If the model is too simple, it is unable to reproduce these important features and extrapolation to the natural referent system becomes risky. If the model is too complex, it becomes as difficult to understand and manipulate as the real thing. Faced with this dilemma, a model builder can improve his model by simplifying it or making it more complicated while retaining consistency.

## REFERENCES


Colby, K.M.
1974. ARTIFICIAL PARANOIA: A Computer Simulation of Paranoid Processes. Pergamon, New York, (In Press).

Colby, K.M., Weber, S., and Hilf, F.O.
1971. Artificial Paranoia. ARTIFICIAL INTELLIGENCE, 2, l-25.

Colby, K.M., Hilf, F.D., Weber, S., and Kraemer, H.
1972. Turing-like Indistinguishability Tests for the Validation of a Computer Simulation of Paranoid Processes. ARTIFICIAL INTELLIGENCE, 3, 199-221.

Colby, K.M. and Hilf, F.O.
1974. Multidemensional Evaluation of a Computer Simutation of Paranoid Thought. In KNOWLEDGE AN0 COGNITION, Gregg, L. (Ed.), Laurence Ehrlbaum and Associates, Potomac, Maryland. (Also appears as Stanford Artificial Intelligence Laboratory Memo AIM-194, Computer Science Department, Stanford University).

Colby, K.M., Parkison, R.C., and Faught, B.
1974. Pattern-matching Rules for the Recognition of Natural' Language Dialogue Expressions. AMERICAN JOURNAL OF COMPUTATIONAL LINGUISTICS. Vol 1, Microfiche 5. (Also appears as Stanford Artificial Intelligence Laboratory Memo AIM-234, Computer Science Department, Stanford University).

Colby, K.M., Hilf, F.D., Wittner, W.K., Faught, B., and Parkison, R.C.
1974. Measuring the Improvment in Linguistic Comprehension in a Model of Paranoid Processes. (Forthcoming).