Eastern Psychological Association Convention
Boston, April 2nd through the 5th, 1992

Development of an Artificial Intelligence Based Role-Play System for Psychological Assessment

Michael L. Raulin and **Michael Anbar**

State University of New York at Buffalo

This report describes a three-year project to develop computerized role-play scenarios to be used in psychological assessment. These programmed scenarios utilize an artificial intelligence program (*CASIP, Computer Assisted Socratic Instruction Program*; Anbar, 1990). Five scenarios were developed and refined to the point that they provide realistic interchanges between the computer and the subject. The subject responds to the computer using natural language input and the computer responds in kind after parsing and then interpreting each subject's response.

The concept of personality has a long and somewhat contentious history in psychology (Lanning, 1991; Mischel, 1968). Personality should theoretically relate to behavior. However, traditional personality measures have often shown disappointingly small correlations with non-test behavior (Mischel, 1984). It is almost a cliche in psychology that the best predictor of future behavior is past behavior in a similar situation. Behavior therapists, relying on this principle, have emphasized the use of response-based assessment of personality, obtaining behavioral samples through role-playing (Barlow, 1981). A disadvantage of role-playing over traditional personality assessment techniques is the professional time required to set up, score, and interpret the subjects' performance in the roles.

The rapid development of computer technology and artificial intelligence concepts (Wagman, 1991) provides us with alternatives to some of the traditional assessment techniques. As the speed, memory capacity, and sophistication of the algorithms improve, the computer is capable of doing much more than simply scoring an objective psychological test or presenting fixed sequences of stimuli to which the subject responds in a multiple choice manner. The study described here uses an artificial intelligence based capability to correctly interpret and respond to natural language input as the basis for a unique approach to personality assessment.

Method

Development Process

Five scenarios were developed as part of a project to improve the evaluation of medical student applicants. The scenarios, which are programmed as a series of linked tables, are implemented using a driver program developed by one of the authors (*CASIP*, Anbar, 1990). The program searches for key words and phrases and their synonyms, evaluating the items found and their relationship to one another in order to determine an appropriate response. The *CASIP* program permits the logical analysis to extend beyond the immediate subject response by remembering and including in the evaluation elements from past responses of the subject. The result is a life-like interaction between a person and a machine, with the exception that the interaction is restricted to a keyboard entry of the person's responses.

The five scenarios were designed to provide challenging social interactions (see Table 1). In each case the computer is playing the role of a particular person whose responses might be benign, bureaucratic, defensive, or accusatory. The subject's task is to recognize the nature of the interaction and respond in an appropriate manner. The consistent personality of each scenario comes across well to subjects who typically rate the scenarios as being very realistic.

Table 1 *Brief Descriptions of the Five Role-Play Scenarios Used in the Study*

---

(1) (COUNSELOR) The first scenario required the subject to assume a somewhat higher social status than the computer-emulated person. Here is the introduction to the scenario: *"You are a student peer counselor. Your role is to advise students on subjects concerning their health and welfare. John, a student, is sent to you by the resident advisor because several student have been complaining about his smoking. What will you say to John?"* John has an inconsistent personality that oscillates between militancy and compliance, making his handling intentionally difficult and frustrating. There is a twist when John mentions, when appropriately interrogated, that the complaint of his mates against him has an ulterior motive--he thinks he is hated because he is a better student and is more popular with girls than his peers.

(2) (LIBRARIAN) This scenario puts the testee in a confrontational position with an unyielding and bureaucratic librarian. This is how it opens: *"You try to check our from the library a book that is essential for a term paper due tomorrow, and the librarian will not let you. She insists that you have an overdue fine of $16.25 because a previous book was returned late. You know that you returned that book on time and that the library is at fault. This library does not issue receipts when books are returned and paying the fine is regarded as a final settlement; usually there is no way to recover a fine once paid. Books in this library are not stamped with dates of check out and check in. What will you say to the librarian at this point?"* The librarian is consistently authoritative and unyielding to a point that evokes frustration.

(3) (LOST WALLET) This scenario puts the testee in an equal status to the emulated person under conditions that may call for suspicion. *"You are living in a dormitory. You just noticed that your wallet containing your monthly allowance, credit cared, and driver's license is missing. There is a young man named Bob in the room. He is a high-school classmate of Pat, your roommate. Bob arrived just yesterday. Pat is going to be in class for the rest of the afternoon. There is a phone on the desk, so you may call Campus Security. What will you say now to Bob?"* This scenario ends with a twist: *"It is 7PM. Pat returns at last and says, while in the doorway, 'I found your wallet in the hallway near the elevator. Since I was rushing to class I could not get back and tell you. Here it is. You must have been worried, weren't you? Well let's go have dinner. By the way, where is Bob?'"*

(4) (GUARD) In this scenario the computer emulates an authoritative bully: *"You buy a pack of pencils in a local drug store. Just as you are leaving the store, a security guard stops you. He accuses you of stealing a pack of gum. The gum was at your feet, just as you were leaving the checkout counter. It must have fallen out of someone else's bag, but to the guard it appears that it fell out of your bag. What will you way now to the guard in your defense?"* Like in the library scenario, the subject is innocent, but this offense is more serious and the consequences of conceding are much more severe. The guard is assertive and stubborn and cannot be talked out of his accusation.

(5) (PARTY) In this scenario the computer emulated an irrational scenario of equal status. *"You are at a Delta Tau Delta fraternity party, and one of the brothers, with the smell of beer on his breath, grabs you and says, 'Hey . . . This is my shirt! Give me it! Right now!' He is very insistent that you have his shirt. Surely you know he is wrong. You are separated from your friends and have no extra clothing. How will you talk your way out of such a situation?"* The accuser is obviously drunk and possibly assaultive. If you give in and offer your shirt, he claims you are wearing his shoes as well and he wants them, too.

By the time the program has analyzed the input sufficiently to respond in an appropriate manner, it also has information sufficient to score the response in a variety of dimensions. We developed a scoring manual focusing on five characteristics that we felt would be adequately sampled in the scenarios. These

characteristics include social skills, combativeness, submissiveness, frustration level, and negotiative skills. Each response is automatically scored by the program according to criteria in this manual, with summary scores available on demand. In addition, the computer records process variables such as the time a subject takes before beginning each response (labeled think time), the time between the last character of a response and the carriage return (reread time), and the number of backstrokes. These variables may reflect a subjects' confidence level.

The development process was iterative, with the programming of the scenarios refined on the basis of repeated testing. Over 200 subjects were tested during a two-year development period until virtually any reasonable subject response was recognized and appropriately responded to by the program. In the data below, almost 90% of subject responses were recognized, with the computer giving appropriate responses 98% of the time. (The program has a series of default responses when recognition fails.)

Initial Validation

The scenarios were given to 71 first year medical students during their orientation to medical school. They were tested on IBM 286 and 386 machines in supervised groups.

Results

The average intercorrelation of our five personality dimensions across scenarios were uniformly low (ranging from -.06 to +.08), suggesting that none of these variables were showing trait-like stability across situations. A *post hoc* evaluation of the automatic scoring feature of the program suggested that there were a few small scoring errors (i.e., the program misinterpreted the nature of the input and assigned an incorrect score), but these errors were rare and were unlikely to have attenuated the correlations much. A more reasonable interpretation of these data is that the scenarios were sufficiently distinct that trait-like consistency in behavior across scenarios is unlikely (Mischel,. 1983, 1984; Shoda, Mischel, & Wright, 1989). Greater consistency across scenarios was observed for the process variables (think time, reread time, proportion of backstrokes), with inter-scenario correlations averaging between .25 and .30. However, the clinical significance of these process variables has yet to be established.

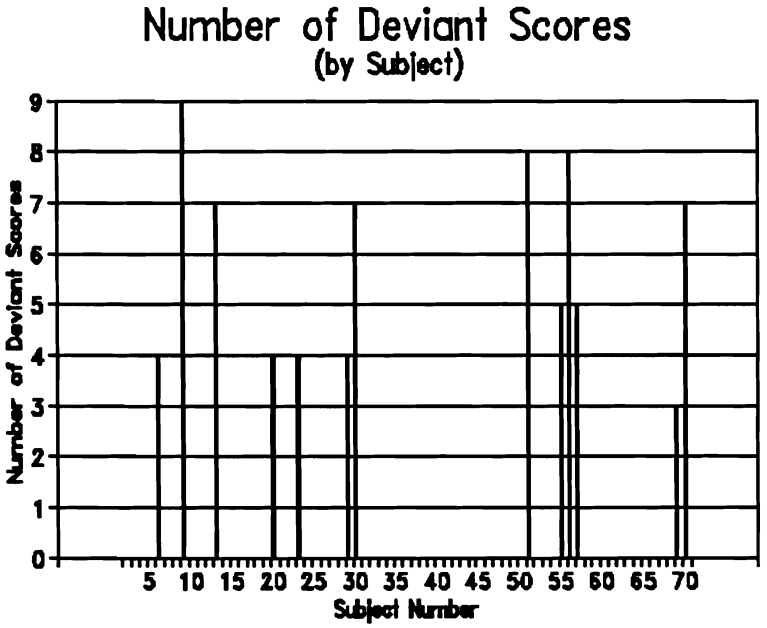

Figure 1 *Distribution by subject number of the number of deviant scores.*

Our most promising analysis focused on outliers on each variable. Frequency distributions usually identified between 1 and 4 people who were clearly deviant on that variable compared to the rest of the sample. Compiling who these deviant subjects were across all of the variables for the five scenarios revealed a surprising degree of clustering, with 13 of 71 subjects accounting for all 72 deviant scores (see Figure 1). The nature of the deviance differed from scenario to scenario, but certain subjects were consistently deviant. Since we do not yet have criterion measures available (i.e., how our subjects do in medical school), we are not in a position to evaluate the clinical significance of this finding. A clinical inspection of these deviant subject's responses, however, suggests that this method of scoring really does identify socially inappropriate behavior.

Discussion

The current data demonstrate the feasibility of a computer generated role-play assessment with automatic scoring of the subject's responses. This is only the first step in a planned series of studies of application of this technology to psychological assessment and treatment. Already in the planning stage is a Socratic instruction program for assertiveness using short role-play scenarios. Since the computer will be evaluating the subjects' responses on-line, individualized instructions can be provided between scenarios on how to improve one's assertiveness technique. The application of this computer technology to behavioral assessment and treatment is almost limitless.

References

Anbar, M. (1990, June). Computer Assisted Socratic Instruction Program (CASIP): Preliminary Documentation. Unpublished manual, available from author (Michael Anbar, Department of Biophysical Sciences, School of Medicine and Biomedical Sciences, State University of New York at Buffalo, Buffalo, NY 14226).

Barlow, D. H. (1981). *Behavioral assessment of adult disorders*. New York: Guilford.

Lanning, K. D. (1991). *Consistency, scalability, and personality assessment*. New York: Springer-Verlag.

Mischel, W. (1968). *Personality and assessment*. New York: Wiley.

Mischel, W. (1983). Alternatives in the pursuit of the predictability and consistency of persons: Stable data that yield unstable interpretations. *Journal of Personality, 51*, 578-604.

Mischel, W. (1984). Convergences and challenges in the search for consistency. *American Psychologist, 39*, 351-364.

Shoda, Y., Mischel, W., & Wright, J. C. (1989). Intuitive interactionism in person perception: Effects of situation-behavior relations on dispositional judgments. *Journal of Personality and Social Psychology, 56*, 41-53.

Wagman, M. (1991). *Artificial Intelligence and human cognition: A theoretical intercorrelation of two realms of intellect*. New York: Praeger.

Eastern Psychological Association Convention
Boston, Massachusetts
April 4, 1992