The Detection of Simulated Malingering Using a Computerized Category Classification Test

Hasker P. Davis, John H. King, Michelle R. Bloodworth, Angela Spring, and Kelli J. Klebe

University of Colorado at Colorado Springs

A category classification test was used to differentiate between normal student control participants, students instructed to malinger a memory deficit, and amnesic patients. Controls (N = 44) and amnesic patients (N = 10) were instructed to do their best, while simulators of malingering (N = 43) were instructed to fake a memory deficit for credit and possible financial compensation. Participants studied a list of high distortions of a prototype dot pattern and were then asked to choose whether or not a new set of dot patterns (random patterns, high distortions, low distortions, and the prototype) belonged to the same category of dot patterns as studied. Malingerers performed significantly worse than normal controls and amnesic patients. A discriminant function analysis showed that the classification test can be used to correctly classify participants as simulated malingerers, controls, or amnesic patients significantly higher than chance. These results indicate that a category classification test can be used in the detection of simulated malingering and that some tests of implicit memory provide a potential supplement to standard forced choice tests in the detection of malingering. © 1997 National Academy of Neuropsychology

THE DETECTION OF SIMULATED MALINGERING USING A COMPUTERIZED CATEGORY CLASSIFICATION TEST

Neuropsychologists are frequently asked to assess individuals reporting emotional and/or cognitive impairments following head injury. The most common cognitive complaint after head injury is memory impairment. In such cases where there is a possibility of malingering, the typical assessment has involved the administration of tests of explicit memory. Such tests require the conscious recollection of information that has a personal and/or temporal context. Among the tests used to detect malingering of memory impairment are the Wechsler Adult Intelligence Scale–Revised (Rawling & Brooks, 1990), Wechsler Memory Scale (Rawling & Brooks, 1990), Portland Digit Recognition Test (Binder, 1993), Rey Auditory Verbal Learning Test (Bernard, 1990, 1991), and the Rey 15-item visual memory test (Bernard & Fowler, 1990). Of these tests, the most commonly used is the forced choice recognition test (Binder, 1993).

We thank Larry R. Squire and Barbara Knowlton for generously providing the task software, the data from amnesic patients reported in this study, and for helpful comments on the design of this study.

Address correspondence to: Hasker P. Davis at the Department of Psychology, University of Colorado, 1420 Austin Bluffs Parkway, P.O. Box 7150, Colorado Springs, CO 80933-7150.
1990; Frederick & Foster, 1991; Hiscock & Hiscock, 1989; Pankratz, 1983). On a forced choice recognition test clients choose between two possibilities, thus making chance performance approximately 50% correct. A score significantly below chance is indicative of malingering (Binder, 1990; Pankratz, 1983). However, the majority of suspected malingerers will not score below chance (Binder, 1993). Accordingly, empirically based “cut-off scores” are also frequently used to distinguish malingerers from individuals having genuine memory deficits. When this method is used, the achievement of a score below the cut-off, but not significantly below chance, provides a basis for suspecting malingering.

The “deficit” demonstrated by the malingerer’s score on the forced choice test, as well as other explicit memory tests, is in the same direction as that of a genuinely memory impaired patient, albeit more severe than would be expected. This overplaying of the memory deficit on the part of malingerers reflects their misconception of how a truly memory impaired individual would perform. Recently, further advantage was taken of the layperson’s misconception about the nature of the memory impairment that characterizes the amnesic syndrome by administering controls and simulators of a memory impairment a test of implicit memory—a domain of memory that does not depend on conscious awareness and includes such phenomena as skill learning, habituation, classical conditioning, and priming (Davis, King, Klebe, Bajszar, Bloodworth, & Wallick, in press). Specifically, participants were administered two word stem priming tests. Participants were induced to study a set of words without being told that later they would be presented with a word stem and asked to respond with the first word that popped into mind. Simulators completed significantly fewer word stems and demonstrated selectively longer response latencies for words previously presented than control participants. In contrast, amnesic patients perform normally on this task when they are instructed to complete word stems with the first word that pops into mind, yet are severely impaired when asked to recall the words previously studied (Graf, Squire, & Mandler, 1984).

Few studies have addressed the potential of implicit tests as an assessment technique for detection of malingering, probably because in previous studies mixed results were obtained. For example, Williamsen, Johnson, and Erikson (1965) reported normal priming in participants with posthypnotic amnesia, but impaired priming in simulators of posthypnotic amnesia. In contrast, Wiggins and Brandt (1988) found that simulators of global amnesia and control participants exhibited similar priming effects on an immediate word stem completion task. This finding seems somewhat surprising because simulated malingerers would not be expected to know that amnesiacs perform at normal levels on priming tasks, and therefore, the simulated malingerers should perform more poorly than control participants. Simulators have an incorrect perception about how to perform on implicit memory tasks, and error by performing poorly. Thus, simulators show a misconception of how an amnesic patient will perform on an implicit memory test, and indeed, perform poorly on a test where normal performance would be expected for a memory impaired individual (Davis et al., in press). Accordingly, they might overplay their role to such an extent that they would perform poorly on a variety of implicit memory tests where normal performance by amnesic patients would be expected. To examine this idea and extend the investigation of malingering to another implicit memory test, the performance of simulators of a memory deficit, control participants, and amnesic patients was examined on a category learning test.

The category classification test used in this study was originally reported by Knowlton and Squire (1993). It was found that amnesic patients and control participants performed similarly when asked to classify novel dot patterns as to whether or not they belonged to the same category as a set of training patterns. This finding and other reports of normal category classification learning by amnesic patients (Kolodny, 1994) provide support for the view that category classification is an example of implicit memory that is independent of explicit memory. Accordingly, it is predicted that simulators of a memory deficit will perceive the
category classification test as a test of memory, and overplay their role of a memory impaired individual by performing significantly worse than control participants and amnesic patients.

METHOD

Participants

Eighty-seven undergraduate psychology student volunteers at the University of Colorado at Colorado Springs were assigned to either a simulated malingering group (n = 43, females = 35, males = 8) or a non-malingering control group (n = 44, females = 33, males = 11). The control group averaged 27.6 ± 10.1 years of age and 14.5 ± 1.3 years of education. Participants in the control group were told that they would be completing a battery of tests that investigated memory and that they should do their best. The simulated malingering group averaged 24.9 ± 7.2 years of age and 14.5 ± 1.1 years of education. There were no significant differences at the .05 significance level between the controls and simulators in age, F(1, 85) = 2.1, p = .15, nor in years of education, F(1, 85) < 1.0, p = .89. The participants in the simulated malingering group read the following instructions:

We would like you to imagine you were in an auto accident where the other driver was at fault. During the accident you struck your head and experienced a significant period of unconsciousness. As a result of the head injury you were hospitalized for 1 week. It has now been several months since the accident and you are having memory problems. You are now involved in a lawsuit to determine the amount of financial compensation to be awarded to you as the victim of the accident. We will be giving you several neuropsychological tests in order to assess your memory problem. Your task is to simulate a memory problem without making it obvious to the experimenter that you are doing so intentionally. Finally, a reward of 50 dollars will be given to the two individuals who simulate the most believable memory problem in the judgment of the experimenters. Do you have any questions? Remember, you will be trying to simulate a believable memory problem on the various tests you will be given. Finally, it is important that you do not tell the examiner the nature of these instructions.

The 10 amnesic patients (three females and seven males) were tested at the Veterans Administration Medical Center in San Diego (Knowlton & Squire, 1993). The participants averaged 65.3 years of age, attained an average of 13.8 years of education, and achieved a mean full scale IQ of 106.3 on the Wechsler Adult Intelligence Scale-Revised. The mean scores on the five indices of the Wechsler Memory Scale-Revised were the following: Attention-Concentration—102.6, Verbal Memory—74.1, Nonverbal Memory—76.9, General Memory—69.1, and Delayed Memory—57.4. Five of the participants (Korsakoff patients) suffered from diencephalic amnesia and five suffered from medial temporal lobe amnesia. Eight of the patients had diencephalic or hippocampal lesions confirmed by quantitative neuroimaging (Shimamura, Jernigan, & Squire, 1988; Squire, Amaral, & Press, 1990), and based on etiology, the two remaining patients were suspected of having hippocampal damage.

Pattern Categorization Test and Procedures

The tests and procedures have been described previously (Knowlton & Squire, 1993). Participants were initially presented a study list of 40 patterns on a computer screen after being read the following instructions:

In the following test you will be seeing a series of dot patterns appearing on the screen. You will use a pencil to point to the dot in the center of the pattern. Each pattern will appear on
the screen for 5 seconds. Please keep your eyes on the screen for the entire time that the dot pattern appears. If the dot pattern leaves the screen before you have pointed to the dot that you think is closest to the center of the pattern, a new pattern will appear and you point to what you think is the dot in the center of the pattern. If you are not sure which of the dots is closest to the center, you should make a guess. Do you have any questions?

For the study list patterns, a prototype dot pattern was constructed by randomly placing nine dots within a 12 × 12 cm area in the center of the computer screen (Knowlton & Squire, 1993). Forty high distortion patterns of this prototype were generated for the study list using the method of Posner, Goldsmith, and Welton (1967). Specifically, for generating distorted patterns, dots were assigned probabilities of moving different distances away from their original position. For high distortion patterns, dots were assigned higher probabilities for movements of greater distance. For low distortion patterns, dots were assigned high probabilities for movements of a short distance, and lower probabilities for movements of greater distance.

After a 5-minute delay, participants were presented with 84 test items for classification. The test items consisted of 4 instances of the prototype, 20 high distortions of the prototype, 20 low distortions of the prototype, and 40 random dot patterns. The random items were constructed using high distortions of new prototypes. No more than three test items of the same type occurred consecutively. Prior to the test, participants were read the following instructions:

The dot patterns that I just showed you a few minutes ago can be thought of as belonging to a single category of dot patterns. In fact, the patterns you saw were actually generated by a set of rules, and were examples of a particular category of dot patterns. Just as an Irish Setter, a Beagle, and a Retriever are all examples of a particular category of dogs, the patterns you saw can also be thought of as examples of a particular type of dot pattern. I’m going to show you some new patterns. These will be ones you have not seen before. Some of these belong to the same category as the patterns in the first part of the experiment, and some do not. If you think the pattern you see belongs in the same category as the patterns you saw in the first part of the test, press the key marked “Yes.” If you think the pattern does not belong in the category, press the key marked “No.” If you are not sure if the pattern belongs in the category, you should make a guess. Do you have any questions?”

The amnesic patients and 36 control participants were administered a second classification test that followed the procedure described above. The test forms were administered in a counter balanced order in separate sessions. The new stimuli were generated in the same manner using different prototypes.

For statistical analysis, the four dependent variables were the percent of prototypes correctly classified, the percent of high distortions correctly classified, the percent of low distortions correctly classified, and the percent of random patterns incorrectly classified.

Shipley Institute of Living Scale (Zachary, 1988)

Control participants and participants simulating a memory deficit were administered the Shipley after completing the pattern classification task(s). All participants were told several times to do their best and ignore any instructions on faking a memory problem if their original instructions told them to do so. The Shipley provides age estimated WAIS-R IQ scores.
TABLE 1
Means and Standard Deviations for Category Classification Test Patterns

<table>
<thead>
<tr>
<th>Condition</th>
<th>Pattern Controls</th>
<th>Simulators</th>
<th>Amnesiacs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protype</td>
<td>80.11 (25.62)</td>
<td>53.49 (31.61)</td>
<td>77.50 (18.45)</td>
</tr>
<tr>
<td>Low Distortion</td>
<td>66.36 (17.33)</td>
<td>51.86 (21.60)</td>
<td>56.00 (11.74)</td>
</tr>
<tr>
<td>High Distortion</td>
<td>66.25 (15.89)</td>
<td>54.19 (18.28)</td>
<td>59.00 (15.77)</td>
</tr>
<tr>
<td>Random</td>
<td>35.69 (18.23)</td>
<td>48.49 (16.67)</td>
<td>47.90 (16.94)</td>
</tr>
</tbody>
</table>

RESULTS

Shipley

A one-way ANOVA was performed on the estimated WAIS-R scores derived from the Shipley Living Institute Scale with group (simulators or control) as the independent variable. Using a .05 significance level, there was no significant difference between the controls (M = 109.2, SD = 5.9) and the simulators (M = 108.0, SD = 6.6), F(1, 85) < 1.0, p = .38.

Category Classification Test

For the amnesic patients and control participants who took two forms of the category classification test, separate dependent t-tests were performed on the percent of patterns correctly classified with form of the test as the independent variable. No significant differences at a significance level of .05 were found for the amnesiacs or the controls on the two forms for prototype, low distortion, high distortion, or random patterns, (p values ranged from .18 to .71). Thus, only a single form of the category classification test will be used in all subsequent analyses.

The means and standard deviations achieved on the category classification test by controls, simulators, and amnesic patients are given in Table 1. A three (condition) by four (pattern) ANOVA on percent of patterns assigned to the learned category was performed with condition (control, simulator, or amnesic) as a between subjects factor, and pattern (prototype, high distortion, low distortion, and random dot patterns) as a within subjects factor. All statistical tests were done with a significance level of .05. The Huyn-Feldt correction factor was used for the within-subject factor as the Mauchly sphericity test indicated that the assumption of homogeneity of covariances was not met, W = .46, χ²(5) = 71.43, p < .001. There were significant main effects of condition, F(2, 94) = 7.38, p < .001, and pattern, F(2, 190) = 19.7, p < .001; and a significant interaction between condition and pattern, F(4, 190) = 8.6, p < .001. This interaction is illustrated in Figure 1.

To further investigate the interaction effect, simple effects of condition were examined for each pattern. A pooled error term was used to test the simple effects (MS error = 430.89) and a significance level of .01 was used to control for increasing Type I error rate with multiple follow-up tests. There was a significant effect of condition for the prototype pattern, F(2, 376) = 19.05, p < .01, and for the low distortion pattern, F(2, 376) = 5.414, p < .01. There were no significant simple effects for the high distortion pattern, F(2, 376) = 3.69, p < .05, nor for the random pattern, F(2, 376) = 4.50, p < .025. Tukey’s HSD tests indicated that simulators scored significantly lower than controls and amnesiacs on the prototype pattern, and they scored significantly lower than controls on the low distortion pattern, while amnesiacs and controls were not significantly different on prototype or low distortion patterns.
Category Classification Test

Subsequently, a linear discriminant function analysis was performed on a set of four predictor variables in order to classify participants into either the simulators, controls, or the amnesic group. The predictor variables used were the percent of prototypes correctly classified, percent of low distortions correctly classified, percent of high distortions correctly classified, and the percent of random patterns misclassified. The entire sample was used to estimate the discriminant function since sample sizes would become too small for accurate estimation and classification if the sample was split for cross-validation (Stevens, 1986). An estimate of the classification rate in future samples is obtained using the maximum posterior probability method (Huberty, 1994).

Wilks' method of estimation was used in the discriminant function analysis. All four variables were significant predictors in the discriminant equation. Wilks' $\Lambda = .68$, $F(8, 182) = 4.88$, $p < .001$. The linear discriminant function analysis correctly classified 27 of 44 controls (61.4%), 22 of 43 simulated malingerers (51.2%), and 6 of 10 amnesic patients (60%). This gives an overall correct classification rate of 57%. The estimated correct
classification rates in future populations are 74% for controls, 56% for simulated malingerers, and 42% for amnesics with an overall classification rate of 57% for future samples. This estimated percentage of correct classifications is significantly larger than the percentage of correct classifications expected by chance. Huberty (1994) suggests the use of an index to measure how much better the linear discriminant function is in classifying participants over a method based on chance. Using this index with the percentage correct expected in future samples, the linear discriminant function should result in about 36% fewer classification errors than would be made if classification were done by chance alone.

In addition, a discriminant functional analysis was performed using only prototype pattern because this was the one pattern on which both the amnesiacs and controls differed from the simulators. This linear function analysis correctly classified 23 of 44 controls (52.3%), 28 of 43 simulators (65.1%), and 5 of 10 amnesic patients (50%). This gives an overall correct classification rate of 58%. The estimation of the classification rates in future samples is 46% for controls, 70% for simulators, and 27% for amnesic patients for an overall classification rate of 47%. This is statistically greater than the percentage of correct classifications based on chance alone. The use of this linear discriminant function would reduce the number of incorrect classifications by about 21% more than what would be obtained if chance alone were used to classify individuals.

Because controls and amnesics were hypothesized *a priori* not to be different and there were no significant differences between these groups on any of the factors, it is reasonable to think of the 12 control participants who were classified as amnesic patients to be correctly classified. Similarly, the three amnesic patients who were classified as controls can be thought of as correctly classified. This raises the percent of correctly classified controls to 79.6 and the percent of correctly classified amnesics to 80 and gives an overall correct classification rate of 73% for the discriminant analysis using just the prototype pattern as a predictor.

**DISCUSSION**

The results indicate that the category classification test used in this study can be useful in the detection of simulated malingering. Simulators performed significantly worse than controls and amnesics on the prototype pattern portion of the classification test. This result is consistent with the view that malingerers perceive the category classification test as a test of memory, and accordingly, they inaccurately portray their memory deficit by correctly classifying fewer prototypes than controls or amnesic patients. There were no significant differences between the conditions for high distortion and random patterns, and only one significant difference on the low distortion pattern for controls and simulators. This indicates that the portion of the classification test that distinguishes the simulators is the prototype pattern. This is consistent with the view that the level of difficulty for the other portions of the classification test are such that the simulators cannot easily distinguish if these patterns belong or do not belong to the studied category. It is only the prototype pattern for which the simulators can effectively mangle. This limitation of the classification test requires consideration. The observation that the pattern distortions and random patterns were not useful in distinguishing between groups suggests that the classification test used in this study was too difficult. A test using low distortions of a well known pattern (e.g., a letter or number) would provide simulators a greater opportunity to perform poorly.

The results from the discriminant analysis indicate that the prototype portion of the classification test is useful in the detection of simulated malingering. That is, up to 79.6% of the control participants, 80% of the amnesic patients, and 65.1% of simulated malingerers
were correctly classified using the predictors employed. Even when corrected for use in future samples the classification rates obtained from using the classification test are significantly higher than the chance rate of 33%.

The results presented in this study provide further support for the use of implicit tests in the detection of malingering. These findings compliment previous studies showing that implicit tests of priming can detect participants simulating memory problems (Davis et al., in press; Williamsen et al., 1965). Further, performance on this test not only distinguished the simulators from a control group, but also distinguished them from a true amnesic group. Caution is still necessary in considering the generalizability of these results to a clinical population of true malingerers because such a group has not been tested. Nevertheless, the present study supports the view that implicit tests of memory have considerable potential as part of a battery of tests for the detection of malingering.

REFERENCES


