BRIEF REPORTS

Chance Guessing in a Forced-Choice Recognition Task and the Detection of Malingering

Kenneth A. Flowers, Carol Bolton, and Nicola Brindle
University of Hull

Guessing occurs on forced-choice (FC) tests for which responses cannot be based on relevant knowledge. Its importance is in inverse relation to the level of knowledge being measured, so that it becomes an increasing component of test scores as the level of knowledge decreases. It is also used as a benchmark to detect simulated impairment. This investigation examined the role of guessing in a 2-alternative FC face recognition test. Chance groups shown only the test items were asked to guess which were the targets, to measure the variation in scores likely to be found with pure guessing. Controls performed normally, and two simulation groups tried to fake amnesia. Results suggested that simple guessing in the chance group produced variable scores that overlapped both low genuine and “malingering” performance. Low control scores were hidden by the guessing “chance bonus,” which ameliorated the apparent level of decline in memory. Simulators told of their role at the outset (before presentation of the target items) produced more convincing “amnesic” scores than those told only just before the test, who produced the expected below-chance level of score. It is suggested that guessing variability should be taken into account in interpreting FC scores.

Keywords: forced-choice tests, recognition memory, guessing rate, malingering

Forced-choice (FC) tests are often used in neuropsychology testing as, for example, in Raven’s Matrices (Raven, 1958) and part of the Rivermead Behavioral Memory Test (Wilson, Cockburn, & Baddeley, 1991). The main advantage of FC tests is the control of the subjective criterion factor (how willing participants are to volunteer answers or to decide yes in yes–no tasks when they are not sure of them), so that partial knowledge can be incorporated into a standardized response and scoring scale.

Another major advantage that FC tests may be said to have is a control for guessing. In the absence of any meaningful knowledge or memory, participants have to guess between a fixed number of alternatives, and the probability of choosing the correct item by chance can be calculated according to the number of alternatives from which the choice is made. This chance factor gives a minimum score on the test, which is assumed (even if implicitly) to be the same for everyone and which provides a baseline for measured levels of performance.

Chance guessing, however, plays a wider role in FC test performance than simply providing a minimum baseline score. In any FC test on which performance is less than perfect, a guessing component is likely to be part of the score, and the worse the level of real knowledge or ability is, the greater this role is likely to be. And where guessing is a major part of the performance, variability in the hit rate becomes a major possibility, which may distort scores significantly. Curiously, very few tests that use FC techniques make an adjustment for guessing, nor do they consider its role in individual scores even when it must make up a major portion of them, such as in individuals with an impairment for whom scores are likely to be low.

More recently, FC tests and the guessing factor especially have been used as the basis of tests to detect malingering or other functional rather than organic impairment on neuropsychological performance. Examples include Tegner (1988) on hysterical paralysis and Pancratz (1979, 1983, 1988), Binder (1992; Binder & Willis, 1991), and Chouinard and Rouleau (1997) for memory. (Malingering here is defined as a patient deliberately underperforming on tests so as to give a false impression of impairment.) Again, the assumption of these tests is that patients attempting to simulate impairment (amnesia, e.g.) can be caught out because they perform significantly differently from those with genuine impairment by scoring below the postulated chance (minimum guessing) level on a FC test.

A universal assumption in FC tests, and in the “below-chance” rationale for detecting malingering, is that the chance guessing factor in FC tests is constant according to the probability given by the number of alternatives. But variation is an inherent quality in the operation of chance, so the question of how much deviation there is from the postulated hit rate of guessing is an important one for any assessment of FC performance, whether genuine or false. The intention of this investigation, therefore, was to test the rationale of the guessing component of FC performance and to see what kind of allowance should be made for it. The main objectives were to see, first, what the actual hit rate (score) was, and how much variation there was on a clinically used two-alternative FC test when participants are asked to just guess the answers; second,
how much overlap there was with low genuine performance and what effect an adjustment for guessing had on scores; and third, whether those simulating impairment always underscored the statistically expected baseline score so as to reliably give themselves away as underperforming. Also, in genuine performance, do participants without enough knowledge to choose the target correctly really guess at random, or are there other criteria that influence guessing? These questions are important for the interpretation of FC test scores and their use in assessing cognitive deficits and also for test design and the detection of false or exaggerated claims of impairment.

The experimental design involved four groups. A control group performed the tests in the standard manner to give genuine memory scores, which could be adjusted for guessing. A chance group was not shown the target material, but was asked to guess from the test sequence alone which item they thought was the target—that is, they performed entirely by guesswork. Two “malingering” groups were tested in the standard manner, but were told to perform the recognition test as they thought someone with a moderate amnesia would do—that is, they were to try to simulate an amnesic impairment. One of these groups, the simulation-at-presentation (SAP) group, was told at the outset, before they saw the target items, that they would be required to simulate. The other group, the simulation-at-testing (SAT) group, was shown the target material in the normal way but told about simulating only just before the test sequence. The two simulation groups were compared to see whether being warned that one was going to simulate before being shown the original targets made it easier to produce a convincing simulation of amnesia.

The experiment involved a commonly used clinical test of face recognition, the Warrington Recognition Memory Test for Faces, or RMTF (Warrington, 1984), a two-alternative FC test with 50 items. The RMTF was chosen because, with 50 target items of relatively difficult material to recognize, it is likely to produce scores that involve some degree of guessing even on genuine performance, and with a two-alternative FC format giving a 50% predicted chance guessing rate, there is maximum scope for variation in successful guessing around this level. Also, because even genuine performance is likely to be uncertain in places in the test, the simulation groups would be faced with a challenge to set their error rate on items they deliberately choose to fail at what they thought was a realistic rate on top of their genuine error score. That is, they could not simply deny a set percentage of items to produce a chosen level of deficit, because they might not be certain what their genuine level of deficit on the test was to start with. These characteristics were thought to make the malingerling ambience of these groups as similar to that of clinical malingerers as possible.

Materials and Procedure

The RMTF is the visual memory half of the Warrington Recognition Memory Test. The targets are 50 photographs of male faces, presented for 3 s each. Participants are asked to say whether they think each face is pleasant or unpleasant. The immediately following test sequence consists of 50 targets, each paired with a distractor randomly placed in the left or right position, and the task is to choose which face was seen before. The test was presented to the control group and the two malingering groups in the standard manner according to the test manual. The chance group was presented with the test sequence only and asked to guess which face was the target. Scores were the number of targets correctly chosen.

Each participant was tested individually, and the time taken to complete the test was approximately 15 min. After age and gender were recorded, the instructions were read out loud, repeated if necessary, and any questions answered. After testing, the SAT and SAP groups were questioned about the techniques they had used to simulate amnesia. Participants were then debriefed.

Instructions given to the four groups were as follows.

Control group. The control group was given normal instructions:

This is a memory test for faces. I am going to show you this pack of faces one at a time and for each face I want you to say “yes” if you think he looks pleasant and “no” if you think he is not so pleasant. There is no right or wrong answer but I do want you to make a judgment about each face. Here is the first photograph. Does he look pleasant or not so pleasant?

At test, the instructions included, “If you can’t remember or are not sure of any item, then just guess.”

Chance group. The chance group was told, “I am going to show you 50 pairs of photographs of ‘unknown’ males from a clinically used recognition test for faces. Please choose from each pair the face which you think is more likely to be the ‘target’ face in the test.”

SAT group (faking at test only). The SAT group received the normal instructions, as above. At test, this group was told,

Now I am going to test your memory for the faces you have just seen in the pack. You are to guess if you are not sure. Before beginning, I would like you to imagine that you are suffering from amnesia (which is a partial or complete loss of memory). Please respond to the test in a way that you believe would convince a psychologist that you are suffering from amnesia. Do you understand? [Pause to check that they do.]

SAP group (faking from the start). The SAP group was told before presentation,

I am about to administer a memory test. Before beginning, I would like you to imagine that you are suffering from amnesia (which is a partial or complete loss of memory). Please respond to the test in a way that you believe would convince a psychologist that you are suffering from amnesia. Do you understand? [Pause to check they do.] This is a memory test for faces [followed by standard instructions as above].

At test this group was told, “Now I am going to test your memory for the faces you have just seen in the pack. You are to
Results

The number of targets (out of 50) correctly chosen by each participant was doubled to give a percentage score. As the main interest of the study was the variability of scores rather than group effects, Figure 1 shows the distribution of scores for all participants in the four groups individually.

Chance Group

The chance group scores center on the 50% level as expected, but the range is from 66% to 32%, which overlaps the control and the malingering groups at either end of the distribution. There is thus some variation in the chance guessing level, which may affect the accuracy of measurement of participants’ recognition levels in both cases. The mean of the chance group is 51.2 (SD = 7.5). Taking the conventional cutoff point for abnormality of 2 standard deviations above or below the mean of a control score, therefore, recognition rates have to be above 65% or below 35% to be outside the domain of guessing. For genuine performance, it implies that a proportion of any score could include a greater chance element than is usually estimated. The effect is not large but may become more important as the level of genuine memory drops and a greater proportion of the test items are included in the guessing component of performance.

For those suspected of malingering, a proportion of their apparently below-chance scores could be due to a greater than expected guessing error, unless the scores are absurdly low. Here a third of the SAT group and half of the SAP group fell above the cutoff point for the minimum chance score and so could not be identified as malingers with any confidence.

Control Group

Scores for the control group are (as in the standard clinical test) not adjusted for guessing, and so are inflated for all but the highest scores, and the range of scores is compressed. FC scores can be adjusted according to a formula first given by Postman (1950): adjusted score = C − W/N − 1, where C = corrected score, W = number of wrong responses, and N = number of alternatives in each test item.

For a two-alternative FC test, this adjustment equates to the number of correct choices minus the number of errors. This means that with a one-in-two FC test, guessing makes up for half the number of any items missed because of a lack of real memory such as that suffered by amnesic patients. The items correctly guessed constitute a “chance bonus” component of the total score, which becomes an increasingly large proportion of the score as the level of real underlying memory decreases. Without the adjustment for guessing, recognition scores on the test are progressively misleading as they drop down the scale. Thus, any score below about 80% (which equates to about 60% real memory level) will show a notable drop when adjusted, with implications for the assessment of the state of underlying memory. About half the scores in the control sample would be markedly reduced in this way.

Simulating Groups

Scores for the two malingering groups show a marked difference in their distribution. The rationale for catching malingers out is based on their tendency to score below chance, which occurs because they fail to understand that on a two-alternative test they are predicted to choose correctly on at least half the items they do not remember (as well as any that they do.) To fake an amnesic impairment, malingers have to deliberately choose the distractor on some items that they recognize, and by definition (as malingers) they know a large proportion of the items. Without adjusting for the chance bonus, therefore, what appears to them to be a moderate degree of “amnesia” (added to the items they genuinely get wrong) means that they in fact choose wrongly on too many items in the test. Here the SAT group showed the expected below-chance tendency, with only 4 participants scoring above the 50% chance level. The SAP group, however, showed much greater variation in scores, from a virtually normal 90% score to a virtually zero score. Nine participants (i.e., about a third) of this group managed to perform at a level that indicates an impaired memory
while still remaining above chance, that is, a good level of faking. It may well be, therefore, that when participants know in advance that they will be trying to simulate amnesia, they can adopt some strategy to improve their faking ability, possibly by not paying attention to the stimuli during the presentation phase. If so, they genuinely will not know what the target is for many items and will, paradoxically, have to choose at random—and so get some right by chance. When the test is presented to them in such a way that they pay close attention to the stimuli during presentation, they are more likely to “not remember” too many items and thus simulate an impaired performance less accurately.

**Guessing and Recognition Rates Across Items**

If the chance guessing group were choosing items at random, one would expect each item to be chosen correctly by half the participants (N = 15) or, given natural variability, for there to be a normal distribution of successful guesses across the 50 items around this mean of 15. The number of times each of the 50 faces was correctly guessed by the chance group was plotted and showed, as expected, a roughly normal distribution centered on the 50% level, with a mean of 15.66 correct guesses and a standard deviation of 3.45. The range was from 9 to 27, but there were a handful of outliers beyond the 1 standard deviation limit (hit rates greater than 19 or less than 12). Guessing in this group does, therefore, appear to be mostly random, with some degree of variation between items. One or two faces were chosen by an abnormally high or low number of participants.

The possibility that the pure guessing rate (as shown by the chance group participants) could affect recognition rate in normal performance was tested by correlating the guessing rate for each item shown by the chance group with the recognition rate for the item by the control group. As this relationship could only show up for participants who have a relatively low score (where guessing is likely to be a major component of their performance), this correlation was calculated for the 14 individuals in the control group who scored 60% or less on the test (i.e., whose underlying memory was calculated to be random guessing will score above the chance level by 10% above the baseline of the 50% guessing rate, but it actually means a loss of 20% rather than 10% (from adjusted score 80% to 60%), and an apparent score of 60% actually means the participant is performing with only 20% real recognition. Conversely, a score of 60% could be misinterpreted as being “only” 10% above the baseline of the 50% guessing rate, but it actually means a genuine recognition score of twice that (20% real memory and 40% guesswork). It does, of course, mean that two thirds of such a score actually consists of the chance bonus, and where variation in this is significant, it means that the accuracy of the apparent score is likely to be compromised. When investigations seek only to show differences between groups or conditions this may not matter much, but it will do so when any scale is used to indicate an actual level of memory or a degree of decline as in dementia.

The two simulation groups differed markedly in their performance on the test. Scores for most of the SAT group showed a reliable below-chance level of recognition, which in the clinic would raise suspicions of malingering, although the scores should really be at least 15%–20% below chance to be really unlikely. The SAP group, however, showed a much greater range of scores, from virtually normal to well below chance, with about a third of the group producing a score similar to that of a genuine amnesia, that is, low but still above chance. This suggests that participants who know that they will be trying to get at least some of the items wrong may adopt some strategy to make it more “genuine,” such as, for example, not looking at the target properly, so that they in fact manufacture a failure to register the items in memory at all. The fact that some participants produce scores in the normal or malingering range shows these efforts may not be very efficient,
but some appear to be able to gauge the effect appropriately. It may be, therefore, that the best way to detect malingerers on FC tests is to force them to pay close attention to the material during presentation in some way so that they always resemble the SAT group during testing. It certainly seems that the higher the level of actual memory malingerers possess (so that chance guessing is a minimal part of their performance), the more likely they are to be caught out by exaggerating their deficit, because they do not seem to take into consideration that on a two-alternative FC test they should be getting somewhat more than half the items correct whatever their impairment. This is, of course, the rationale of tests proposed to detect malingering by below-chance scores, and why they all use only a two-choice format (Chouinard & Rouleau, 1997; Pancratz, 1979, 1988).

The requirement for a significantly worse-than-chance score to detect malingering, however, means that only out-and-out malingerers will be caught by this method. What is perhaps a more common condition, namely patients who have a genuine impairment but want to exaggerate it, may paradoxically be more difficult to detect by this means because their genuine amnesia will show the normal compensation of guessing on items they do not know, which means that they will have a greater chance bonus factor in their performance despite themselves. Hence, they will not perform notably below the minimum chance level and thus their malingering will be less obvious.

References

Received August 31, 2006
Revision received August 10, 2007
Accepted October 15, 2007