An evaluation of two commonly used tests of unfamiliar face recognition

Bradley C. Duchaine a,*, Andrea Weidenfeld b

a Department of Psychology, Vision Sciences Laboratory, Harvard University, 33 Kirkland Street, William James Hall, 7th Floor, Cambridge, MA 02138, USA
b Institut für Psychologie der Universität Potsdam, Potsdam, Germany

Received 19 March 2001; received in revised form 25 September 2002; accepted 25 September 2002

Abstract

The Warrington Recognition Memory for Faces (RMF) and the Benton Facial Recognition Test (BFRT) are commercially available tests that are commonly used by clinicians and cognitive neuropsychologists to evaluate unfamiliar face recognition. Yet, it is not clear that a normal score on either instrument demonstrates normal unfamiliar face recognition. Because the RMF's stimuli contain abundant non-internal facial feature information, subjects may be able to score in the normal range without using internal facial features. On the BFRT, subjects commonly rely on feature matching strategies using the hairline and eyebrows rather than recognizing the facial configuration. To test whether these routes to recognition can support normal performance, normal subjects were tested with versions of the RMF and the BFRT in which the faces had been painted over in a way that prevented the operation of some of the procedures normally involved with face recognition. Even though these modifications removed all of the internal feature information in the RMF, many subjects scored in the normal range, and despite precluding the use of configural processing in the BFRT, many of the scores were in the normal range. As a result, it is apparent that normal scores on these tests do not demonstrate normal unfamiliar face recognition and so clinicians should be cautious in interpreting scores in the normal range. Finally, these results place in question models supported by dissociations involving normal performance on these tests.

© 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Prosopagnosia; Neuropsychological tests; Warrington; Benton; Object recognition

1. Introduction

The Warrington Recognition Memory for Faces (RMF) [51] and the Benton Facial Recognition Test (BFRT) [8] are commercially available tests that are widely used by both clinicians and cognitive neuropsychologists to assess face recognition abilities. However, there is reason to question whether normal scores on these tests actually demonstrate normal unfamiliar face recognition. The photos in the RMF include many non-internal facial features, such as clothing and hair, by which they can be recognized. This raises the possibility that participants—in particular individuals with face recognition impairments—might be able to do well on the RMF by using a strategy that does not require the integrity of normal face recognition processes. The BFRT has the same problem, but for different reasons. Because it asks participants to match faces that are presented simultaneously, it is not, however, show that face recognition processes are intact, because it appears that face recognition is performed by both parts-based procedures that represent facial features and configural processing procedures that represent the spatial relations of the parts of the face [14,25,26,37,40,44]. Moreover and possibly more importantly, target faces and test items in the BFRT are presented simultaneously so participants are not required to rely on a memory trace.

As a result of these concerns about the RMF and BFRT, we conducted tests exploring these possibilities with modified versions. In our version of the RMF, parts-based and configural procedures could not be used on internal feature information, and our version of the BFRT precluded the use of configural procedures. If subjects can achieve a normal score nevertheless, then it will be clear that normal scores do not necessarily demonstrate normal face recognition abilities.

1.1. Recognition Memory for Faces

The RMF was designed as a test of non-verbal memory, but it is often used as a test a particular type of non-verbal
memory—face memory. Subjects in the RMF are presented with black and white photos of 50 unfamiliar men at a rate of one item every 3 s, and they are asked to respond 'yes' or 'no' according to whether they find the face pleasant or not pleasant. Immediately following presentation of the target photos, memory for the photos is tested with a forced choice between a target photo identical to that presented earlier and a distracter photo. This means that stimulus recognition, not true face recognition, is possible [29]. The number of correct choices determines the score of the subject, so the maximum score is 50. Means range between 42.4 and 44.3 depending on the age group with standard deviations of approximately 3.5.

Even though the RMF is used to test face recognition, the photos contain information that is not internal facial feature information and could be used to discriminate between target and distracter photos. When attempting to test face recognition, the distinction between information handled by face recognition procedures and other information is clearly crucial. Although a number of investigations of face recognition have proposed that both the internal features (eyes, nose, mouth regions) and the external features (hair, ears) are recognized by face recognition procedures [21,22,54], neuropsychological evidence paints a different picture. Humans appear to use special procedures for face recognition [11,24,31,37,40], and the input from the face that these procedures typically operate on defines what counts as facial information. CK, an object agnosic with normal face recognition [11,24,31,37,40], and the input from the face that these procedures operate on defines what counts as facial information. CK, an object agnosic with normal face recognition [11,24,31,37,40], and the input from the face that these procedures operate on defines what counts as facial information. CK, an object agnosic with normal face recognition [11,24,31,37,40], and the input from the face that these procedures operate on defines what counts as facial information.

Recent experiments with a developmental prosopagnosic vividly illustrate the problem caused by the presence of the non-internal feature information [42]. EP’s performance was clearly impaired on tests of familiar and unfamiliar face recognition. However, EP was able to achieve a normal score on the RMF, and he reported using the non-internal feature information. In order to investigate his claim, EP was presented with the RMF without the information around the face. Not surprisingly, the mean for the control subjects was only two points lower in the modified version, because they were relying primarily on the internal feature information. In contrast, EP scored 31/50 on the modified test, and so it appears that the non-internal feature information does provide a route to normal performance.

1.2. Benton Facial Recognition Test

Two versions of the BFRT [8] can be used depending upon time considerations: the Short Form has 13 items with 27 possible points while the Long Form has 22 items with 54 possible points. On each item, subjects are presented with a target photo, and they are asked to choose the target individual from six faces presented simultaneously with the target photo (see Fig. 2). There are three parts to the BFRT: (1) matching a frontal view of the target with an
identical photograph, (2) matching a frontal view of the
target individual with three photos of the target taken from
different angles, and (3) matching a frontal view of the
target individual with three photos of the target taken under
different lighting conditions. No time limits are placed on
the BFRT, and scores are classified as normal when 41 or
above.

The black and white photos used in the BFRT consist of
unfamiliar male and female faces with their hair and cloth-
ing shaded out so that subjects must rely on the face. On
many of the items, the features that best differentiate the pho-
tographed individuals are the eyebrows and hairlines: normal
subjects and prosopagnosic subjects to whom we have ad-
ministered the test, as well as prosopagnosics in [20,41,42],
have reported that their performance relied heavily on match-
ing the eyebrows and hairlines rather than the internal facial
features.

1.3. Assessing the RMF and the BFRT

Because of the possibility that normal scores on the RMF
and the BFRT do not reflect normal face recognition abilities,
we have designed experiments that will determine how well
participants can perform on these tasks without using all
of the processes involved with normal face recognition. For
the RMF, we covered the faces in the test phase so that
participants were forced to rely on only the non-internal
feature information available in the photos (see Fig. 3). For
the BFRT, we covered all of the face except for the eyebrows
and the hairline, so that participants could not use the facial
configurations for matching (see Fig. 4). We will refer to
these modified versions of the task as the mRMF and the
mBFRT.

Because we deleted internal feature information, this al-
lows us to test normal subjects under conditions similar to
those experienced by individuals with normal parts-based
processing but without facial configural processing. How-
ever, we deleted internal features that could have been
used by parts-based processing so our versions provide an
especially demanding test of the hypothesis that these tests
can be passed using feature-based procedures. If some
participants are able to perform in the normal range on
either test, it will indicate that the original test does not
require normal face recognition abilities. However, if par-
ticipants have scores well out of the normal range, it will
indicate that the original tests do, in fact, require intact face
recognition.

2. Experiment 1a: modified Recognition
Memory for Faces (mRMF)

2.1. Method

2.1.1. Participants

The participants were 26 undergraduate students, 17
women and 9 men, who participated to satisfy a require-
ment for their introductory psychology class at UC-Santa
Barbara.

2.1.2. Materials

The stimuli were digitized versions of the stimuli from
the Warrington Recognition Memory for Faces test. The 50
two-choice test stimuli were scanned into a computer so that
the faces could be painted over with a shade of gray. The gray
extended from each model’s hairline to his chin vertically
and horizontally from ear to ear or from sideburn to sideburn,
if sideburns were present (see Fig. 3). We then copied the
target stimulus from each of the forced-choice pairs for use
in the study phase of the experiment. This insured that each
target stimulus shown during the study phase was identical
to the target stimulus used in the test phase. Digitized stimuli
were the same size as the stimuli in the RMF, and each
stimulus pair in the test phase was presented side by side
with the same spacing used in the normal RMF.
2.1.3. Procedure

Participants were seated at computers and were presented with instructions on the screen. The instructions informed them that they would view people with their faces painted over, but pointed out that there was a large amount of non-internal feature information available to utilize for the recognition test that would follow. Participants were encouraged to use any mnemonic strategies they felt comfortable with, and it was suggested that giving the target individuals names or noting imperfections in the photos might be effective methods.\(^1\)

Following the instructions, participants were presented with the 50 target stimuli for 3 s each. A white screen was presented for 1 s between each target. At the end of the study phase, participants read instructions directing them to right click on their mouse if the target was on the right and to left click if the target was on the left. If they were unsure which stimulus was the target, they were instructed to guess. As in the normal RMF, reaction time was not measured. However, we required that the participants make their decision within 10 s. After reading these instructions, participants responded to the 50 test pairs.

2.2. Results

Because the participants were college age, the scores were compared to the 18–39-year-old sample provided in the RMF manual. Scores of 38 or above are classified as normal, and 17 of the 26 participants would be classified as normal on the basis of their scores on the mRMF. In other words, in the absence of internal feature information, 65% of the participants were able to achieve a score that would label them as unimpaired on unfamiliar face recognition. The average for the participants was 38.5 with a standard deviation of 4.04.

3. Experiment 1b: modified Recognition Memory for Faces (mRMF)

The previous experiment showed that the non-internal feature information alone provides enough information for participants to score in the normal range. However, the procedure differed in a few respects from the standard RMF, and these differences may have inflated the scores on the mRMF. First, participants were not required to make a ‘pleasant/unpleasant’ decision with regard to each face. Although past experiments indicate that such a decision produces better face recognition performance than an intentional memory strategy [10,52], it may have raised scores in our test. Second, we presented stimuli in the study period that only had non-internal feature information; of course, the RMF study stimuli had internal facial features and non-internal feature information. Because faces are stimuli that are especially likely to capture attention [34,49,50], omission of the internal feature information may have made the non-internal feature information easier to attend to and remember than it would be in the standard RMF.

In order to determine if these differences inflated the scores of the participants, we have modified our procedure with the mRMF in experiment 1b. Participants are required to make a ‘pleasant/unpleasant decision’, and the study stimuli are the normal study stimuli that have both internal features and the non-internal feature information. If participants are still able to score in the normal range, this will reinforce our previous results and eliminate the concerns discussed above.

3.1. Method

3.1.1. Participants

The participants were 23 undergraduate and graduate students, 10 males and 13 females, from Harvard University who participated for a small payment.

3.1.2. Materials

The test stimuli in this version of the mRMF were the same as the previous experiment. The study stimuli were created by copying the target stimulus prior to painting over the internal feature information from each test pair. This assured that the study and test versions of the target were identical other than the deletion of the internal feature information in the test stimuli.

3.1.3. Procedure

The procedure was identical to the previous procedure except that participants were required to make a ‘pleasant/unpleasant decision’ after viewing each study stimulus. Participants indicated their decision with one of two keyboard responses. As in the previous experiment, the study stimulus was presented for 3 s. A white screen was presented immediately following each study stimulus, and it remained until participants made the ‘pleasant/unpleasant’ decision after which the next study stimulus appeared.

3.2. Results

Scores of 38 or above are considered in the normal range, and eight of the 23 participants scored in the normal range. This proportion—35%—was less than that seen in experiment 1a, but the non-internal feature information still allowed more than one-third of the participants to achieve a normal score. The average score was 35.4 with a standard deviation of 4.68.

\(^1\) Our instructions varied in a number of respects from the regular instructions, but we did this to encourage subjects to use strategies that individuals with face recognition impairments use to compensate for their deficits.
4. Experiment 2: modified Benton Facial Recognition Test (mBFRT)

4.1. Method

4.1.1. Participants

The participants were 29 undergraduate students, 10 males and 19 females, from UC-Santa Barbara who participated to satisfy a requirement for their introductory psychology course.

4.1.2. Materials

In order to eliminate the possibility of configural face processing, the stimuli from the BFRT were modified to leave only the eyebrows and the hairline visible. After scanning the original test items, gray masks were painted over the face of each model up to the eyebrows (see Fig. 4). The modified stimuli from each item on the BFRT were then placed on a common canvas in the same configuration as stimuli in the BFRT: the target stimulus was placed above two rows of test stimuli. As in the BFRT, numbers were placed below each test stimulus so that subjects could easily identify each stimulus.

4.1.3. Procedure

Participants were seated at a computer, and read the instructions for the task on the monitor. They were informed that for the first six problems there was only one correct answer and that the remaining problems had three correct answers. Participants were provided with an answer sheet on which they were to circle the stimulus numbers that they perceived to match the target stimuli. Like participants in the standard BFRT, they were allowed to proceed at their own pace.

4.2. Results

One point is awarded for every correctly matched test stimulus. There are six possible points on the first six items and 48 possible points for the remaining 16 items for a total of 54 points. The Short Form consists of the first 13 items, and it has a total of 27 possible points. The Long Form consists of all 22 items, and so there are 54 possible points. Because participants responded to every item on the mBFRT, each subject produced a score for the first six items, a Short Form score, and a Long Form score. For the first six items, the target and the matching test item were identical images, and the participants’ average of 5.85 showed that they had very little trouble with these items. On the Short Form, participants averaged 20.41 points with a standard deviation of 2.15. A Short Form score of 20 converts to 41 on the Long Form, and 41 is classified as a normal score.

Inspection of individual scores shows that the performance on the Long Form, and 41 is classified as a normal score. On the Short Form, participants averaged of 16 of the 27 subjects on the Short Form would be categorized as normal. Therefore, with only the hairline and eyebrows, 59% of the participants would be classified as normal using the Short Form and 41% would be considered normal using the Long Form.

After the first six problems, test images differed either in the illumination of the faces or in the orientation of the faces. There were eight items of each type, and scores were higher for the orientation changes than for the lighting changes. The mean for the sum of the orientation change items was 18.89 with a standard deviation of 1.95 while it was 15.30 with a standard deviation for the 2.23 for the lighting changes.

5. Discussion

In order to assess the RMF and the BFRT, we tested participants with modified versions of each test. Test faces in the mRMF were painted over so that they provided only non-internal feature information; the stimuli in the mBFRT presented only the eyebrows and the hairline. The mBFRT did not allow either parts-based procedures or configural processing to operate on internal feature information in the test phase, and the mBFRT did not permit configural processing. Despite this, many participants scored in the normal range on both tests. It bears noting that individuals with face recognition impairments are forced to rely on alternative routes to face recognition, and so can be expected to be particularly adept at using them for recognition. In contrast, the participants in this experiment presumably depend on these other cues less often. Thus, one might expect prosopagnosics without other perceptual impairments to perform better on the RMF and the BFRT than the normal participants did on the mRMF and the mBFRT. The results, which show that normal scores on these tests do not demonstrate normal unfamiliar face recognition, have two major implications. First, neither the RMF nor the BFRT, at least in their present form, should be used to diagnose unfamiliar face recognition without additional tests. Second, it is necessary to reconsider aspects of theories of normal face recognition for which performance on the RMF or the BFRT served as critical evidence.

5.1. Recognition Memory for Faces

The results with the mRMF clearly showed that the non-internal feature information was rich enough to support scores in the normal range. Although scores were lower in the second experiment with the mRMF, many participants were still able to score in the normal range. In addition, because our test stimuli did not have internal features, there were fewer features for feature-based procedures to operate on. Our results then raise the question of whether subjects actually do, in fact, use this information. The performance of the prosopagnosic tested by Nunn et al. [42] clearly showed that individuals sometimes do use this information.
This report and our results suggest that normal scores achieved with the RMF may not, in fact, indicate intact face recognition abilities. However, it does not cast doubt on testing done with the RMF showing impaired performance, because individuals with impaired scores were unable to perform normally with both the internal features and the non-internal feature information present.

5.2. Benton Facial Recognition Test

Our results with the mBFRT showed that the presence of only the eyebrows and the hairline were sufficient to support normal performance. Subjects were able to match these features in a piecemeal fashion by comparing the simultaneously presented target photograph and test photographs. Because our stimuli did not permit the operation of configural face processing, a normal score on the BFRJT does not demonstrate normal unfamiliar face recognition abilities. This conclusion is also supported by reports of prosopagnosics with impaired unfamiliar face recognition who have scored normally on the BFRJT [20,41,42]. These individuals have typically reported that they performed the task with the same feature matching strategy that our normal participants were forced to use in the mBFRT. Of course, our results show that normal scores on this test may not reflect intact face recognition processes. However, as was the case with the RMF, impaired performance does indicate impaired face recognition and so results out of the normal range on the test are informative.

Despite the problems that we have found with the BFRJT, there are changes that could be made to improve it as a test unfamiliar face perception. When prosopagnosics have achieved normal scores, the investigators have often reported that they took a long time to complete the test [20,41,42]. This suggests that reliance on a feature matching strategy requires more time, and so it may be possible to test configural face perception by adding time norms or by limiting the time allowed for each item. In order to probe unfamiliar face recognition rather than face perception, tests should incorporate these considerations in a modified version of the BFRJT which allow feature matching or included significant non-internal feature information. In addition, response time data were never collected in the experiments supporting the dissociation, and so it is possible that abnormally long response times may have allowed impaired participants to achieve normal accuracy scores [27,55].

Therefore, in every case appearing to show evidence for normal performance on unfamiliar face recognition tasks with impaired performance on familiar face recognition, the evidence is open to question. Until subjects are tested using better designed tests of unfamiliar face recognition and found to show the dissociation, there is no unequivocal support for this half of the purported double dissociation between unfamiliar and familiar face recognition exists. Theories proposing the existence of a module for each task were driven by the apparent double dissociation between the tasks, so when this dissociation is undermined it is unnecessary to propose multiple modules. In order to incorporate these considerations in a modified version of the Bruce and Young [11] framework, we suggest that the face recognition units recognize both familiar faces and unfamiliar faces.

Acknowledgements

Advice and encouragement were provided by Leda Cosmides, John Tooby, and Ken Nakayama, and we also wish to thank Elizabeth Levy, Cathy Galvez, and Crystal McMillan for their assistance in constructing and running these
experiments. These studies were supported by the James S. McDonnell Foundation, the NSF (Grant BNS 9157-499), a UCSB Research Across Disciplines Grant, and the NIH (Grant 1 F32 MH64246-01).

References


