CYBERPSYCHOLOGY & BEHAVIOR
Volume 7, Number 1, 2004

Spectacles, Distinctiveness, and Face Recognition: A Web-Based Experiment

MITCHELL METZGER, Ph.D.,[1] and K. ROBERT BRIDGES, Ph.D.[2]

ABSTRACT

This study sought to clarify previous research in the face recognition literature regarding memory for faces with spectacles. A second aim of this research was to further investigate Valentine's face-space model, a leading model of face recognition that predicts better performance on distinctive faces compared to typical faces. Prior to this experiment, independent observers provided distinctiveness ratings for faces with, and without, spectacles. Experimental participants then accessed the PsychExperiments website and completed a face recognition experiment. Based on the judgments of the independent observers, the face-space model predicts that memory for spectacled faces should be superior to memory for non-spectacled faces. An analysis of hit rate (percent correct) supported this notion, as a higher hit rate was observed for spectacled faces compared to non-spectacled faces. However, the analysis of false alarms (false identifications) did not support the predictions made by the face-space model, as participants demonstrated reliably higher false alarm rates for faces with spectacles. A further analysis of response bias suggests that the overall pattern of responding may have been largely due to changes in response criteria for trials where spectacled faces were presented. Implications for models of face recognition are discussed.

INTRODUCTION

IN RECENT YEARS, there has been an abundance of research on face recognition, with many investigations of "distinctive" and "typical" faces. Numerous researchers have reported that participants show superior performance on faces previously rated as distinctive compared to faces that are typical in appearance.[1–3] This robust finding has been observed across a wide-variety of testing procedures and conditions, and several theories have been proposed to explain this distinctive face advantage. One leading theory is the face-space model,[2] which has received strong experimental support and appears to account for many of the phenomenon associated with distinctiveness. In essence, the model proposes that faces are composed of many different dimensions (hair length and color, nose size, face shape), and that faces are normally distributed across these various dimensions.[2] Typical faces, as they are more common, are located near the central tendency of face space where there are an abundance of faces, while distinctive faces (those that are less common) are located in areas of face space that have lower population densities. The model posits that when

[1]Department of Psychology, Ashland University, Ashland, Ohio.
[2]Department of Psychology, The Pennsylvania State University, New Kensington, Pennsylvania.

participants perform a face recognition task, they attempt to match the target face with a representation of that face in memory. When matching typical faces, confusion (error) is more likely to occur because there are many other faces located in that area of face space. That is, since many typical faces are clustered around the central tendency of face space, the result is a greater number of faces to "choose from," and errors are, therefore, more likely. Distinctive faces, on the other hand, are better remembered because they are farther from the central tendency of face space (sparser areas), resulting in fewer faces to choose from, which makes confusion less likely.[4] The face-space model has been used to make accurate predictions that not only account for the effects of distinctiveness, but also the effects of inversion and the other-race bias,[2] decision latency effects,[5] and caricature effects.[6]

In an earlier series of experiments on memory for typical and unusual (i.e., distinctive) faces, researchers observed that faces with certain features, such as eyeglasses, were rated as less typical than faces without spectacles.[7] If faces with eyeglasses are less typical, this would imply that they are more distinctive (or, in the words of these researchers, more unusual), which should make them easier to identify in a recognition test. This is precisely what was observed,[7] a finding consistent with the face-space model.[2] However, this was a post-hoc observation, and more recent evidence from studies that have more systematically manipulated the presence of eyeglasses contradicted these early findings. For instance, two reports have shown that poorer performance is observed on trials that present spectacled faces.[8,9] Specifically, these reports have shown that this effect occurs because participants exhibit significantly more false alarms (false identifications) on trials depicting faces with spectacles.

If faces with spectacles are less typical,[7] then performance for faces with eyeglasses should be superior to performance for faces without spectacles, according to the face-space model.[2] However, other findings clearly contradict this notion.[8,9] While each of the aforementioned experiments investigated the presence of spectacles for face stimuli, none of these studies directly manipulated the presence of eyeglasses and its effect on distinctiveness. That is, it is possible that the spectacled faces used by one researcher[8,9] were less distinctive than those used by the other,[7] thus producing the differences in performance scores between the studies.

The purpose of this study was to further clarify the relationship between spectacles, distinctiveness, and face memory. Specifically, this study was designed with two aims in mind. First, we sought to rectify the differential results thus reported on memory for spectacled faces.[7–9] Secondly, we wanted to test whether the face-space model[2] could accurately predict performance on a face recognition procedure where the presence of eyeglasses was systematically manipulated. While the face-space model was not initially formulated to predict a subjects' performance for faces with spectacles, it would be interesting to determine if this leading theory of face recognition would be able to account for face stimuli that have this common characteristic.

MATERIALS AND METHODS

Participants

Data for the face recognition procedure were collected from 110 participants who logged on to PsychExperiments (http://psychexps.olemiss.edu) over a three month period (10/10/02 through 1/13/03). The participants consisted of 64 females and 46 males who ranged in age from 18 to 47, with an average age of 22.57 years. Of those who identified their race, 88% were Caucasian, 5% were African-American, and 3% were Asian-American (4% of participants declined to answer this question).

Materials

The faces used in the face recognition experiment consisted of 48 black and white photographs (head and shoulders) of females taken from a college yearbook that were scanned and converted to jpeg images. Twenty-four of these photographs depicted females with eyeglasses, and 24 depicted females without eyeglasses. Care was used to eliminate photographs of people with distinctive makeup, jewelry, clothing, or other distracting stimuli that would draw attention away from the face. For the distinctiveness rating procedure, the faces were incorporated into Microsoft PowerPoint and projected onto a screen in a room with elevated seating. For the face recognition experiment, the stimuli were incorporated into Macromedia's Authorware 5.1, and this file was housed at the PsychExperiments server at the University of Mississippi. Participants accessed the PsychExperiments website to access the face recognition experiment and the face stimuli were presented onto the user's PC monitor. When the stimuli were projected on the computer screen, they were approximately 5.75 cm tall × 3.81 cm wide (as measured on a monitor with 15″ viewable area). (Participants completing the experiment on computers with monitors other than 15″ would be presented with slightly larger or smaller face stimuli than those reported here.)

Design and procedure

Prior to the face recognition experiment, 85 independent observers (undergraduate students enrolled in Introductory Psychology at the Pennsylvania State University) rated approximately one hundred black and white photographs of female faces on a 9-point distinctiveness scale, ranging from 1 = very typical, to 9 = very distinctive. Each face appeared for 5 sec with a 2-sec blank screen between each face, and the observers made their responses on a prepared answer sheet. The data from the observers were compiled, and an average distinctiveness score for each face was computed. From these ratings, the 24 spectacled faces with the highest scores and the 24 non-spectacled faces with the lowest scores were selected. The faces with spectacles received an average rating of 5.38, while faces without eyeglasses received an average rating of 3.89. This difference was confirmed with a one-way ANOVA, as the faces with eyeglasses were rated as significantly more distinctive than faces without spectacles [$F_{1,46} = 39.4$, $p < 0.001$]. Thus, based on the observer ratings, the faces with spectacles will now be referred to as the "distinctive" faces, and the faces without eyeglasses will now be referred to as the "typical" faces.

Participants for the face recognition experiment (different from those who completed the rating portion of the experiment) accessed the PsychExperiments website (http://psychexps.olemiss.edu) to complete the study. Before the experiment began, they first read an informed consent document and were required to click on the *"Yes, I wish to participate"* button to indicate their consent. If a participant instead chose the *"No Thanks"* button, the program was terminated and they were directed back to the home page for PsychExperiments. If the user agreed to participate, they were assigned a random identification number (i.e., FMB679). As each participant was identified by this random number, it ensured that their data would be anonymous to anyone who accessed the data file. In addition, the participants were asked to provide several pieces of demographic information (age, gender, and race). Once the demographic information was entered (participants were required to enter a value for the gender and age question, but were permitted to choose *"I decline to answer"* on the question of race), the instructions for the experiment were presented on the computer screen. The instructions read by the participants were as follows, and remained on the screen until the *"continue"* button was clicked:

> During the first phase of this experiment, 24 faces will be individually presented on the center of your computer screen. Each face will appear for 2 seconds, and there will be a 1 second interval between each face, during which time the screen will be blank. During this portion of the experiment, you are to examine each face carefully, as a recognition test will follow the initial presentation of the faces. When you are ready to begin the experiment, please click the continue button and the experiment will start.

Once the continue button was clicked the study phase began and the 24 faces were presented. Out of the original 48 faces in the database, the program randomly selected 24 faces to present during the study phase. One-half of the faces depicted females with eyeglasses, and one-half of the faces depicted females without eyeglasses (the stimuli chosen and the order of presentation during the study phase was randomized for each participant). Immediately following the study phase, participants were required to click on another *"continue"* button to proceed to the recognition test. During the recognition test, all 48 faces were randomly presented, one at a time, on the center of the computer screen. Twenty-four of the faces were targets (previously seen during study), and 24 were distractors (new stimuli that were not viewed during study). Just as during the study phase, one-half of the distractor faces depicted faces with eyeglasses and one-half depicted faces without eyeglasses. As each face appeared on the screen, the text *"was this one of the faces you saw previously?"* appeared at the bottom of the screen. Above the text were two clickable buttons, one labeled *"yes"* and one labeled *"no"*. The face remained on the screen until the participant made a response, which was followed by a 1-sec blank screen before the presentation of the next face. As in the study phase, the presentation of faces during the recognition test was randomized for each participant. As each response was made, the program identified the response as a hit, false alarm, correct rejection, or a miss, and compiled these scores for each participant.

After the completion of the experiment, participants were prompted to send the data from their session to the PsychExperiments server, where the demographic and performance data could be downloaded by the experimenter. For each participant, the dependent variables of hit rate and false alarm rate were compiled for faces with spectacles and for faces without spectacles. If a participant completed the experiment and did not wish to submit their data, they clicked on the *"do not send"* button and their data were deleted. Finally, participants read a debriefing statement prior to the termination of the experiment. Once the experiment was terminated, the user was directed to the homepage for PsychExperiments.[10,11]

RESULTS

Mean hit rate (HR), false alarm rate (FAR), and response bias (c) for spectacled and non-spectacled faces can be seen in Table 1. Participants showed higher HR for faces with eyeglasses compared to faces without eyeglasses. This impression was confirmed with statistical analysis, as an ANOVA yielded a significant difference between the two types of face stimuli ($F_{1,109} = 10.95$, $p < 0.001$). This result is predicted by the face-space model; performance on distinctive faces (in this case, those with eyeglasses) was superior to performance on typical (non-spectacled) faces. However, this pattern of results (superior performance for spectacled faces) was not obtained for FAR. Participants made more false alarms on faces with eyeglasses compared to faces without spectacles. This impression was also confirmed statistically, as the ANOVA yielded a significant difference ($F_{1,109} = 14.13$, $p < 0.001$). This result is opposite to what the face-space model would predict, as the model posits lower false alarm scores for distinctive faces. However, this result of higher FAR scores for distinctive faces does replicate earlier findings.[8,9]

In this experiment, distinctive faces (those with eyeglasses) produced reliably higher HR and FAR scores than typical faces (those without eyeglasses). While these data confirm some but contradict other predictions made by the face-space model,[2] this pattern of responding is what one would expect if participants exhibited different response criteria on judging spectacled and non-spectacled faces. That is, if the presence of eyeglasses made the spectacled faces "more similar" to each other (because of the common feature of spectacles), it may make participants more likely to respond *"yes, I have seen this face before"* on trials where spectacled faces were presented. The face-space model hypothesizes that distinctive faces are located in an area of face space with lower population densities, thus making confusion with other faces less likely to occur. While faces with spectacles might be more distinctive-looking than faces without eyeglasses (and, therefore, place spectacled faces in areas of face space with lower population densities), the fact that spectacled faces share this common feature of might increase the population density of that area of face space. This would lead to confusion with other spectacled faces, and may make participants more willing to respond *"yes, I have seen this face before."* If this interpretation is accurate it would, therefore, produce higher HR and FAR scores on trials where spectacled faces are presented.

To determine if this was a plausible interpretation of the data, an analysis of response bias (c) was computed, a statistic that determines a participant's willingness to respond *"Yes, I have seen this face before"* or *"No, I have not seen this face before"* on each trial. For the c statistic, scores near 0.00 indicate no bias, positive scores indicate a bias to respond that a face is new (or, in this experiment, to respond *"no, I have not seen this face before"*), and negative scores indicate a bias to respond that a face is old (or, in this experiment, to respond *"yes, I have seen this face before"*).[12] If the presence of spectacles changed response bias in our experiment (in the sense that a participant would be more likely to respond "yes" on trials with eyeglasses), one would predict significantly negative c scores for trials with spectacled faces, and little, or no, change in c for non-spectacled faces. Table 1 clearly shows that participants exhibited negative c scores for spectacled faces and, to a smaller degree, positive c scores for non-spectacled faces. Further analysis demonstrated this to be a significant negative bias for spectacled faces, and no significant change from 0.00 for non-spectacled faces. That is, response bias for spectacled faces was significantly different from 0.00 ($t_{109} = 4.00$, $p < 0.001$), but response bias for non-spectacled faces did not significantly differ from 0.00 ($t_{109} = 1.71$, $p > 0.05$). Thus, in relation to predictions made by the face-space model, it appears that the contradictory findings that were ob-

TABLE 1. MEAN HIT RATE (HR), FALSE ALARM RATE (FAR), AND RESPONSE BIAS (C) FOR SPECTACLED AND NON-SPECTACLED FACES.

	Spectacled	*Non-spectacled*
HR	0.832 (0.161)	0.768 (0.180)
FAR	0.236 (0.190)	0.170 (0.143)
c	−0.211 (0.554)	0.091 (0.556)

Standard deviations are presented in parentheses.

tained in this experiment were largely due to changes in the criteria for responding on trials with spectacled and non-spectacled faces.

DISCUSSION

It should be noted that there are limitations to these data. Namely, there is the possibility that the spectacled faces were more similar to each other than the non-spectacled faces, which could also cause response bias to change between faces with and without eyeglasses. Further research investigating this phenomenon may manipulate the presence of eyewear in a more systematic way. That is, this could be controlled by using the same faces with and without spectacles as stimuli. In any event, no matter what the reasons are for the observed changes in response bias, this study clearly demonstrates the importance of measuring response bias in face recognition studies.

These data support some, yet contradict other predictions made by the face-space model.[2] While the face-space model can accurately account for many phenomena observed in face recognition studies, care should be taken when interpreting data from stimuli with unusual characteristics, such as eyeglasses. As noted earlier, the face-space model was not specifically designed to account for faces with spectacles. However, spectacles can be a distinguishing feature of some faces, just as can a large nose, bushy eyebrows, or a long chin, and models of face recognition should acknowledge these defining characteristics of certain face stimuli. These data imply that faces with other types of distinguishing features (e.g., scars or facial hair) might also be subject to the same types of results. That is, performance on these other types of faces may produce higher HR and FAR scores, which, at least from the data collected in this study, are largely reflective of overall changes in response criteria for faces with the distinguishing feature of spectacles.

ACKNOWLEDGMENTS

We thank Ken McGraw, Mark Tew, and John Williams for the excellent training in Authorware and for hosting this experiment at the PsychExperiments website.

REFERENCES

1. Johnston, R.A., & Ellis, H.D. (1995). Age effects in the processing of typical and distinctive faces. *The Quarterly Journal of Experimental Psychology* 48A:447–465.
2. Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology* 43:161–204.
3. Vokey, J.R., & Read, J.D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory and Cognition* 20:291–302.
4. Valentine, T. (2001). Face space models of face recognition. In: Wenger, M.J., Townsend, J.T., eds., *Computational, geometric, and process perspectives on facial cognition: contexts and challenges.* Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 83–113.
5. Valentine, T., & Bruce, V. (1986). The effects of distinctiveness in recognizing and classifying faces. *Perception* 15:525–535.
6. Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: testing the face-space framework. *Psychological Science* 11:379–385.
7. Light, L.L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory* 5:212–228.
8. McKelvie, S. (1987). Recognition memory for faces with and without spectacles. *Perceptual and Motor Skills* 65:705–706.
9. McKelvie, S. (1988). The role of spectacles in facial memory: a replication and extension. *Perceptual and Motor Skills* 66:651–658.
10. McGraw, K., Tew, M., & Williams, J. (2002). The integrity of web-delivered experiments: can you trust the data? *Psychological Science* 11:502–506.
11. Metzger, M.M. (2003). The world wide web and the laboratory: a comparison using face recognition. *CyberPsychology & Behavior* 6:613–622.
12. MacMillan, N.A., & Creelman, C.D. (1991). *Detection theory: a user's guide.* New York: Cambridge University Press.

Address reprint requests to:
Dr. Mitchell Metzger
Department of Psychology
Ashland University
401 College Ave.
Ashland, OH 44805

E-mail: mmetzger@ashland.edu