Vision Research 129 (2016) 69–76

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Gender differences in recognition of toy faces suggest a contribution of experience

Kaitlin F. Ryan, Isabel Gauthier *

Department of Psychology, Vanderbilt University, Nashville, TN 37240, USA

ARTICLE INFO

Article history:
Received 15 September 2016
Accepted 11 October 2016

Keywords:
Face recognition
Experience
Learning
Gender differences

ABSTRACT

When there is a gender effect, women perform better then men in face recognition tasks. Prior work has not documented a male advantage on a face recognition task, suggesting that women may outperform men at face recognition generally either due to evolutionary reasons or the influence of social roles. Here, we question the idea that women excel at all face recognition and provide a proof of concept based on a face category for which men outperform women. We developed a test of face learning to measures individual differences with face categories for which men and women may differ in experience, using the faces of Barbie dolls and of Transformers. The results show a crossover interaction between subject gender and category, where men outperform women with Transformers' faces. We demonstrate that men can outperform women with some categories of faces, suggesting that explanations for a general face recognition advantage for women are in fact not needed.

1. Introduction

Many biases in face recognition have been reported: people are more skilled at recognizing faces within their own race (Lindsay, Jack, & Christian, 1991), age (Yovel et al., 2012), and species (Pascalis, de Haan, & Nelson, 2002). These performance differences for face recognition are typically attributed to underlying differences in experience (Gauthier et al., 2014; Yovel et al., 2012). Gender differences have also been reported, however their interpretation is less clear. Women outperform men on several face recognition tasks (Goldstein & Chance, 1970; Lewin & Herlitz, 2002; Lovén, Svärd, Ebner, Herlitz, & Fischer, 2013; Rehnman & Herlitz, 2007). While there is not always a difference in performance between genders (Duchaine & Nakayama, 2006), interestingly, no advantage for men with male faces has been reported.

There are many possible explanations for the advantage that women show on face recognition tasks. This includes differences in memory and social-cognitive skills rooted in differential brain connectivity between genders (Ingalhalikar et al., 2014), differences in gaze preferences that present from a very young age (Baron-Cohen, 2002), gender differences in social interaction, including a greater concern with the attractiveness of other women in female observers, and other socially-motivated goals (Goldstein & Chance, 1970; Lewin & Herlitz, 2002; Sawada et al., 2014; Yovel et al., 2012). Evolutionary pressures or cultural influences could lead to greater efficiency in encoding, remembering, and labeling faces for women (Wolff, Kemter, Schweinberger, & Wiese, 2014; Lovén et al., 2013) – but importantly, these explanations pertain to face recognition as a whole.

However, if differential experience is an important driver of gender differences in face recognition, it may be possible to find categories of faces for which men outperform women. The goal of this study is not to identify the specific causes of gender biases in face recognition. Instead, we seek evidence for a crossover interaction between observer gender and face categories. Finding such a pattern would rule out any account of gender differences at the level of the entire face domain, thereby constraining theoretical explanations. An analogy in studies of object recognition illustrates the benefits of such an interaction. One study described a male advantage on a test of car recognition and suggested that this could be due to better mental rotation in men then women (Dennett et al., 2012). Even though the result was obtained with cars, the authors considered an explanation that would apply more broadly to all object categories. It is not rare that authors consider object recognition as a unitary skill, such that performance for one category is considered representative. This means that the explanation offered for a male advantage in car recognition becomes a hypothesis for a domain-general advantage where men should outperform women for any object category. However, performance with more object categories was measured in later work, women

* Corresponding author at: 308 Wilson Hall, 111 21st Avenue South, Nashville, TN 37240, USA.

E-mail address: Isabel.gauthier@vanderbilt.edu (I. Gauthier).

http://dx.doi.org/10.1016/j.visres.2016.10.003
0042-6989/

outperformed men with some of these categories (McGugin, Richler, Herzmann, Speegle, & Gauthier, 2012). A crossover interaction reveals that gender effects for objects are domain-specific, likely influenced by experience with various categories, rather than requiring one broad domain-general explanation.

Here, we set out to create reliable tests to measure individual differences in face recognition across different types of faces (in particular, toy faces) for which men and women may differ in experience. We created a task similar to the Vanderbilt Expertise Test (VET; McGugin et al., 2012) and the CFMT (Duchaine & Nakayama, 2006), in which participants learn a set of identities across four face categories (Caucasian female faces, Caucasian male faces, Barbie doll faces, and Transformer action figure faces) and later recognize these identities among distractor faces.

Foreshadowing our results, we found a cross-over interaction in performance for men vs. women with Barbie vs. Transformer faces - most significantly, we found that Transformer faces are one category of faces for which men outperform women.

2. Methods

The tests we describe are a subset (Caucasian female, Caucasian male, Barbie, Transformer) of a new battery of face tests, the Vanderbilt Face Expertise Test (VFET; Ryan & Gauthier, 2014). We also included the VET car sub-test (VET-car; McGugin et al., 2012) as a measure of non-face recognition performance where we expect men to outperform women based on previous findings.

2.1. Participants

297 participants completed the VFET face tasks and the VET-car. Participants completed the tasks either in the lab or via Amazon Mechanical Turk (AMT). AMT is an online crowdsourcing platform that has been used to conduct psychological studies and produces results comparable to those in the laboratory with more diverse samples (see Cho et al., 2015; Crump, McDonnell, & Gureckis, 2013; Richler, Floyd, & Gauthier, 2014). All participants were compensated with course credit or a small payment. Self-reported age, gender, and ethnicity are reported in Table 1. Participants provided informed consent per the guidelines of the Institutional Review Board of Vanderbilt University. All research was carried out in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

2.2. Materials and procedures

All participants (either in the lab or via AMT) completed the VFET face tasks (Caucasian female, Caucasian male, Barbie, Transformer) and the VET-car task.

Table 1
Summary statistics for all participants, separated by sex and source. Parentheses in age indicate standard deviation.

	All Participants	Lab participants	AMT participants
N	295	179	116
Age	27 (10.5)	23 (6.9)	33 (11.8)
Men	161	94	67
Women	134	85	49
Caucasian	196	109	87
Black	43	33	10
Asian	41	31	10
Hispanic	7	1	6
Other	8	5	3

2.2.1. VFET

We collected images of human faces from professional photographers and used them with their permission. We collected toy faces from online collections.

For each category we collected 6 different target identities, each with 7 exemplar images (see Fig. 1). This included sets for 6 male and 6 female individuals (84 images of human faces), 6 different Barbie dolls and 6 different Transformer figures (42 images of Barbie doll faces and 42 images of Transformer faces). Barbie dolls and "Barbie-friend" dolls have been released with several different face sculpts since 1959. A large number of Transformer toys have been released since 1984. Therefore, it was possible for us to collect groups of images that corresponded to unique toy identities, similar those of human faces. Exemplar images, while corresponding to a single identity, could vary across many dimensions such as expression, pose, background context, camera view, and lighting condition. For each category we also collected 102 distractor images of other identities, separate from the 6 target identities. During the tasks, each trial contained 1 exemplar image of a learned target identity and 2 previously unseen distractor images of the same face category.

All images were shown in greyscale and 200×200 pixels. We chose to keep some level of background noise in our images, which generally works better with non-posed photos that vary in viewpoint, and is consistent with a more naturalistic context for face recognition and. We chose matching targets and distractors to minimize the diagnosticity of these cues and applied a Gaussian filter to blur backgrounds and further reduce the information from non-face features (e.g. hair, background cues).

2.2.2. VET-car

Participants also completed a modified version of the VET-car (McGugin et al., 2012; catch trials were added to the original test). Car images consisted of 1997–2003 sedan models commercially available in the United States.

2.2.3. Procedure

Participants first reported their perceived level of experience with each face category before being tested. Participants were asked to rate their experience while considering their "interest in, years of exposure to, knowledge of, and familiarity with each category compared to other individuals." Participants responded using a Likert scale with 1 meaning "very much below average" and 9 meaning "very much above average" (see Gauthier et al., 2014 and VanGulick, McGugin, & Gauthier, 2015, for validation of this measure). One limitation is that participants may vary in the reference group they choose to use.

Next, participants completed the VFET face tasks (Barbie, Caucasian female, Caucasian male, Transformer faces) the VET-car (see Fig. 1). Each task was blocked by category. For each category, participants viewed a 3×2 array of the 6 target identities to study for as long as they needed. Participants then completed 6 "identical" trials with triplets consisting of one exact image of a face identity (or car model) they had previously studied and 2 distractor images of novel identities (or models). The target identity could occur in any of the 3 triplet positions and participants indicated whether the target occurred on the left, middle, or right position. Feedback showed the correct location of the target identity. The study array was presented again, and participants completed another 6 "identical" trials (for a total of 12 "identical" trials).

After viewing the study array a third time, participants completed another 39 trials (36 "transfer" trials plus 3 "catch" trials): These showed a new exemplar of a target identity (under different viewing conditions (e.g., position of the face, lighting) and 2 distractor images. No feedback was given on these trials. The "catch" trials for each category were included to test for understanding of