

Greater visual averaging of face identity for own-gender faces

Jan W. de Fockert · Ben Gautrey

Published online: 25 January 2013
© Psychonomic Society, Inc. 2013

Abstract Recent evidence suggests that observers can rapidly form an average representation based on a set of simultaneously presented faces. Here, we replicate this finding and show that the tendency to process sets of faces in terms of an average representation is greater for own-gender faces. Male and female participants viewed sets of four male or female faces before deciding whether or not a subsequently presented single test face had been present in the set. Incorrect endorsement that it was one of the set members was greater when the test face was a morphed average of the four faces than when it was an actual set member, and this effect was strongest when the gender of the faces was the same as the observer's. The finding that observers were more likely to incorrectly endorse own-gender (vs. other-gender) faces forms an exception to the often reported own-gender advantage in face recognition.

Keywords Face averaging · Face perception and recognition · Perceptual categorization and identification

Introduction

There is growing evidence that sets of similar items can be represented in terms of a summary description, such as mean size (e.g., Ariely 2001; Chong & Treisman, 2005). For example, after briefly seeing a set of differently sized circles, observers are remarkably accurate at estimating the mean size of the set and often mistake a subsequent test circle with the same size as the mean set size as one of the set members (e.g., Ariely, 2001). Visual averaging has also

been shown for images of human faces. When presented with sets of multiple faces differing in identity (De Fockert & Wolfenstein, 2009; Haberman & Whitney, 2009) or emotional expression (Haberman & Whitney, 2007) and asked subsequently to judge whether a single test face had been part of the previously seen set, people mistakenly choose a morphed average of all seen faces more often than they choose an actual set member. The finding that items resembling the average of multiple seen items are likely to be endorsed as set members suggests that the visual system can efficiently represent large amounts of visual information by extracting certain summary statistics from sets of similar items. Further support for the notion that faces may be represented by averaging information from multiple face instances has come from the finding that face recognition is better in systems that represent multiple instances of a face in terms of a morphed average, thereby forming a stable image in which accidental variation between instances is averaged out (e.g., Jenkins & Burton, 2008, 2011).

Since performance on face recognition tasks tends to differ between the sexes (Cross, Cross, & Daly, 1971; Ino, Nakai, Azuma, Kimura, & Fukuyama, 2010; Lewin & Herlitz, 2002; Rehnman & Herlitz, 2006, 2007; Shaw & Skolnick, 1994, 1999; Wright & Sladden, 2003), the aim of present study was to investigate gender differences in the tendency to represent sets of faces in terms of a summary description. There is some indirect evidence to suggest that there may be gender differences in visual averaging. Females tend to process more contextual visual information, as compared with males (Barnett-Cowan, Dyde, Thompson, & Harris, 2010), who are more likely to adopt a local bias in visual processing (Phillips, Chapman, & Berry, 2004). Together with the evidence that distributed attention leads to more accurate mean estimations than does focused attention (Chong & Treisman, 2005), this implies that the more focused male processing style may be associated with a

J. W. de Fockert (✉) · B. Gautrey
Department of Psychology, Goldsmiths,
University of London, New Cross,
London SE14 6NW, UK
e-mail: j.de-fockert@gold.ac.uk

reduced tendency to process multiple items in order to extract a summary representation.

An often reported finding is that females are better than males at recognizing female faces (Cross et al., 1971; Lewin & Herlitz, 2002; Rehnman & Herlitz, 2006, 2007; Shaw & Skolnick, 1994, 1999; Wright & Sladden, 2003), whereas male observers are sometimes better at recognizing male faces than female faces (Ino et al., 2010; Shaw & Skolnick, 1994, 1999; Wright & Sladden, 2003). The own-gender advantage in face recognition is generally explained in terms of differences in interest and motivation (McKelvie, Standing, St. Jean, & Law, 1993) or perceptual expertise (Lovén, Herlitz, & Rehnman, 2011). In the present article, we investigated whether the own-gender advantage would also be evident when observers had to indicate whether a test face had been present in a previously seen set of faces, and the face could be either a true set member or a morphed average of all set members. We argued that an own-gender advantage would lead to greater correct endorsement of set members and reduced incorrect endorsement of morphed average faces, for own-gender (vs. other-gender) faces.

Male and female participants viewed sets of four male or female faces before being asked whether a single test face had been present in the preceding set (as in De Fockert & Wolfenstein, 2009). Test faces could be either a photograph of a real person or a morphed average of the photographs of four people, and the measure of interest was participants' tendency to incorrectly endorse the morphed average as a member of the set. Note that whereas using a morphed average as a representation of multiple seen faces would lead to better performance on tasks requiring matching a single face instance with the stored representation (Jenkins & Burton, 2008), our task was specifically designed so that an averaging strategy would lead to a performance impairment. In other words, the task encouraged processing the individual set members, and we argued that any evidence that morphed averages were mistaken for set members would provide compelling evidence for visual averaging of faces.

We tested two predictions regarding possible gender differences in performance on the task. The first prediction was based on the greater local processing bias reported for males (Barnett-Cowan et al., 2010; Phillips et al., 2004). The prediction was that if females have a greater tendency to represent the four set faces in terms of an average description, they should be more likely than males to incorrectly endorse the morphed average of a previously seen set of faces. The second prediction was based on the own-gender advantage reported in face recognition studies (e.g., Cross et al., 1971). On this account, we predicted that all observers would show greater correct endorsement of own-gender (vs. other-gender) repeated member faces. Conversely, all observers should show greater incorrect endorsement of other-gender (vs. own-gender) morphed average faces. A

further basis for the second prediction was the out-group homogeneity effect (Park & Judd, 1990; Quattrone & Jones, 1980), which suggests that out-groups are seen as less variable than in-groups. A key function of perceptual averaging is to create a single representation of a category of similar members. Items must therefore be classified as members of the same category in order to be summarized to form a single representation of the category. Crucially, if the out-group is seen as a more homogeneous category than the in-group, information about members of out-group categories may be more likely to be averaged together than information about in-group categories, which are seen as more heterogeneous. Other-gender faces are thus perceived as one category more than own-gender faces and would, therefore, be more likely to be represented in terms of a summary description. On our task, this would mean that other-gender faces will be more likely to be incorrectly endorsed than own-gender faces, so that incorrect endorsement of a morphed average face, as compared with a repeated photograph, would be greater for male faces in female observers and for female faces in male observers.

Method

Participants

We tested 30 volunteers, between 19 and 40 years of age, 15 males and 15 females. Participants were tested individually and had normal or corrected-to-normal vision.

Apparatus, stimuli, and procedure

The experimental protocol was identical to one we reported previously (De Fockert & Wolfenstein, 2009), and we provide only a brief description of the methods here. We used E-Prime software (Schneider, Eschman, & Zuccolotto, 2002) to present sets of male or female faces (2,000-ms duration), followed by a single male or female test face (presented until response), which could be one of the faces that were present in the previous set (matching photograph), a morphed average of the four preceding set faces (matching average), a face that was not present in the previous set (nonmatching photograph), or a morphed average of four faces other than the ones in the preceding set (nonmatching average). Participants were instructed to make a fast and accurate keypress response to indicate whether or not they thought that the test face had been present in the preceding set. Half the trials contained male, the other half female face stimuli, and within each trial, only faces from a single gender were used (see Fig. 1 for example stimuli). Each participant first completed a practice block of 32 trials, followed by four experimental blocks of 60 trials each.

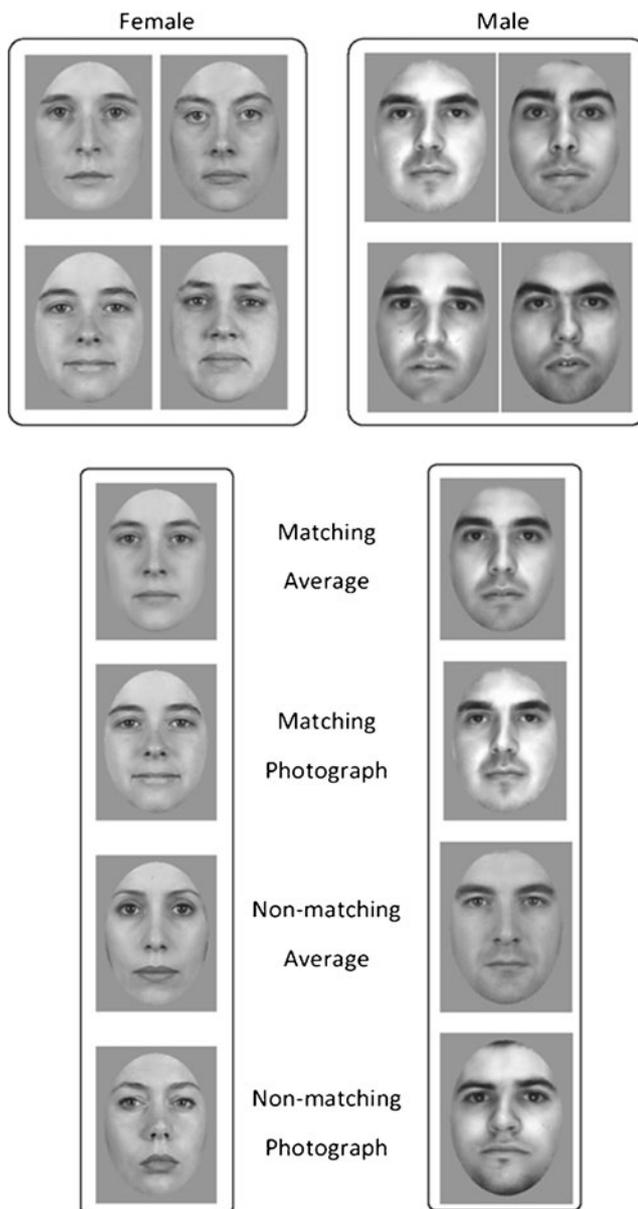


Fig. 1 Example female (*left*) and male (*right*) face stimuli. At the top are sets of four faces. Below are the four possible test faces, with, from top to bottom, the morphed average of the presented set, a photograph from the presented set, the average of another set, and a photograph from another set

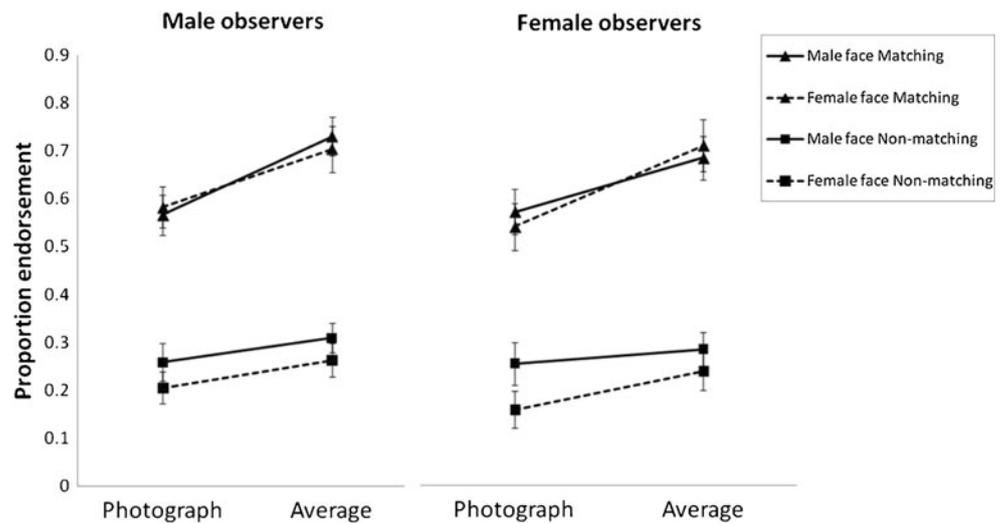
Results

For each participant, the percentage of *present* responses was calculated per condition and was first entered into a 2 (male observer vs. female observer) \times 2 (male test face vs. female test face) \times 2 (average test face vs. photograph test face) \times 2 (matching test face vs. nonmatching test face) mixed analysis of variance (ANOVA) with observer gender as the between-subjects factor. The mean proportions of *present* responses per condition are presented in Fig. 2.

All three within-subjects main effects were significant. Matching test faces were more likely to be endorsed (mean proportion of *present* responses = .599, $SE = .026$) than were nonmatching test faces ($M = .234$, $SE = .022$), $F(1, 28) = 208.84$, $MSE = .038$, $p < .001$, $\eta_p^2 = .882$. Average test faces were more likely to be endorsed ($M = .462$, $SE = .023$) than were photograph test faces ($M = .370$, $SE = .022$), $F(1, 28) = 25.56$, $MSE = .020$, $p < .001$, $\eta_p^2 = .477$. Finally, male test faces were more likely to be endorsed ($M = .431$, $SE = .022$) than were female test faces ($M = .401$, $SE = .021$), $F(1, 28) = 7.03$, $MSE = .008$, $p < .02$, $\eta_p^2 = .201$. Overall, endorsement rates were higher in male observers ($M = .453$, $SE = .029$) than in female observers ($M = .379$, $SE = .029$), although the between-subjects main effect of observer gender did not reach significance, $F(1, 28) = 3.31$, $MSE = .101$, $p = .08$, $\eta_p^2 = .106$.

There were three significant interactions. First, there was a significant two-way interaction between type of test face (average vs. photograph) and matching (matching vs. nonmatching test face), $F(1, 28) = 13.37$, $MSE = .008$, $p < .01$, $\eta_p^2 = .323$. When the test faces matched the preceding set, an average test face was more likely to be endorsed ($M = .666$) than a photograph test face ($M = .532$), $t(29) = 5.56$, $SEM = .024$, $p < .001$, Cohen's $d = 0.85$. When the test did not match the preceding set, an average test face was also slightly more likely to be endorsed ($M = .259$) than a photograph test face ($M = .208$), $t(29) = 2.9$, $SEM = .018$, $p < .01$, Cohen's $d = 0.41$. The interaction shows that the difference between the average and the photograph was greater for the matching than for the nonmatching faces. This effect replicates our previous finding that a test face that is the average of the seen set faces is more likely to be endorsed than an actual member of the set (De Fockert & Wolfenstein, 2009). Second, there was a significant two-way interaction between test face gender (male vs. female) and matching (matching vs. nonmatching test face), $F(1, 28) = 6.57$, $MSE = .006$, $p < .025$, $\eta_p^2 = .190$. For female test faces, the difference between matching ($M = .597$) and nonmatching ($M = .206$) test faces was somewhat greater than for male test faces ($M = .601$ vs. $M = .262$ for matching and nonmatching member test faces, respectively). In other words, whereas correct endorsement of matching test faces differed little between male and female faces, male test faces were incorrectly endorsed more often than female test faces. Finally, there was a significant three-way interaction between face gender, type of test face, and observer gender, $F(1, 28) = 5.02$, $MSE = .003$, $p < .05$, $\eta_p^2 = .152$. For female observers, the difference between endorsing mean versus member faces was greater for female faces (mean difference between mean and member endorsement = .110) than for male faces (mean difference between mean and

Fig. 2 Mean endorsement rates as a function of observer gender, face gender, morphing, and matching. Error bars represent between-subjects standard errors



member endorsement = .064). By contrast, for male observers, the difference between endorsing mean versus member faces was somewhat greater for male faces (mean difference between mean and member endorsement = .108) than for female faces (mean difference between mean and member endorsement = .090). No other interactions were significant (all p s > .3).

The key effect of interest was the gender difference in the tendency to endorse average and photograph own-gender and other-gender faces when they matched the preceding set. The nonmatching conditions, in contrast, were added primarily to ensure that the test face was associated with the preceding set on half of the trials (the matching conditions), but not on the other half (the nonmatching conditions). In order to test any differences in endorsement rates in matching conditions only, we performed a three-way ANOVA (observer gender \times test face gender \times test face) on these data. There was a main effect of test face, $F(1, 28) = 29.96$, $MSE = .018$, $p < .001$, $\eta_p^2 = .517$, again confirming that endorsement rates were higher for matching morphed average faces than for matching photographs. The critical finding, however, was a significant three-way interaction between observer gender, test face gender, and test face, $F(1, 28) = 4.28$, $MSE = .004$, $p < .05$, $\eta_p^2 = .133$. As can be seen in the matching conditions in Fig. 2, male and female observers showed opposite patterns of endorsement rates as a function of face gender. Males had a larger increase in endorsement rate for male morphs (vs. photographs; endorsement difference between morph and photograph of .164), $t(14) = 4.11$, $SEM = .04$, $p = .001$, Cohen's $d = 0.73$, as compared with female morphs (vs. photographs; endorsement difference between morph and photograph of .122), $t(14) = 2.79$, $SEM = .044$, $p < .05$, Cohen's $d = 0.49$. Conversely, females had a larger increase in endorsement rate for female

morphs (vs. photographs; endorsement difference between morph and photograph of .149), $t(14) = 4.43$, $SEM = .034$, $p < .001$, Cohen's $d = 0.59$, as compared with male morphs (vs. photographs; endorsement difference between morph and photograph of .1), $t(14) = 3.01$, $SEM = .033$, $p < .01$, Cohen's $d = 0.44$. Thus, although endorsement rates were reliably greater for matching morphs than for matching photographs in all cases, the interaction confirms that the differences were greater for male faces in male observers and for female faces in female observers.

Discussion

Observers were more likely to mistake a morphed average of four faces previously seen in a set as one of the set members than they were to correctly identify a repeated photograph as a set member. This finding replicates previous evidence suggesting that multiple faces can be represented in terms of a summary description (De Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007, 2009) and provides further support for the notion that information from multiple faces can be summarized to form stable face representations. Clearly, this mechanism applies not only to repeated instances of the same facial identity (Jenkins & Burton, 2011), but also to sets of faces with multiple identities. Moreover, averaging does not necessarily require sequential processing of instances of the same facial identity (the prototype effect; e.g., Bruce, Doyle, Dench, & Burton, 1991; Cabeza, Bruce, Kato, & Oda, 1999) but also occurs for simultaneously presented faces that were perceived just once. Finding evidence for visual averaging in a task in which this strategy impairs performance and in which, instead, the importance of processing of

individual set members is emphasized provides strong evidence that visual averaging is a fundamental mechanism for face processing.¹

The key finding was that the likelihood of choosing the morphed average face over a photograph was greater for male faces in male observers and for female faces in female observers. This pattern of the gender effect, with stronger averaging for faces with the same gender as the observer's, was not one that we predicted. Our first prediction, that females would show a greater tendency to endorse the morphed average of a set of seen faces than would males, was not supported: The three-way interaction between observer gender, face matching, and face condition (average vs. photograph) was not significant, $F < 1$, implying that female observers were as likely as male observers to endorse matching morphed average faces. Our second prediction, based on both the own-gender advantage in face recognition (e.g., Cross et al., 1971) and the out-group homogeneity effect (e.g., Park & Judd, 1990), was that incorrect endorsement rates for own-gender average faces would be lower and correct endorsement rates for own-gender photographs higher, as compared with faces from the opposite gender. Instead, we found the opposite pattern of results, with higher incorrect endorsement of own-gender (vs. other-gender) morphs and lower correct endorsement of own-gender (vs. other-gender) photographs.

Our findings are not in line with the prediction, based on the own-group heterogeneity effect (Park & Judd, 1990; Quattrone & Jones, 1980), that averaging would be stronger for items that were members of an out-group category. This could mean that visual averaging occurs relatively independently of perceived set similarity. However, previous work has shown that the perceived similarity of items within a group does affect averaging, to the extent that people derive separate summary statistics from multiple sets of similar items that were nonetheless presented simultaneously (Chong & Treisman, 2005). When presented with displays of differently sized red and green circles, observers subsequently are shown to have knowledge of the different mean sizes of the red and green subsets, rather than of the set as a

whole. It thus may be that gender is insufficiently perceived as a factor producing in-group and out-group identifications to affect visual averaging. Other categorizations, such as ethnic group or age group, may produce a stronger perception of “us” and “them,” and such factors may well show the effects on averaging we predicted on this view. Further research is needed to address this issue.

The present findings also form an exception to the often reported own-gender advantage in face recognition (e.g., Cross et al., 1971; Lewin & Herlitz, 2002; Rehnman & Herlitz, 2006, 2007; Shaw & Skolnick, 1994, 1999; Wright & Sladden, 2003). We should first note that our task was designed to measure people's tendency to rely on summary representations in face processing, rather than face recognition per se, and, as such, was different from the tasks used in previous work on the own-gender bias, such as tasks for face recognition memory (e.g., Cross et al., 1971) or simultaneous face-matching tasks (e.g., Megreya, Bindemann, & Havard, 2011). Apart from methodological factors, what could explain the greater averaging for own-gender faces? A speculative possibility is that any greater perceptual expertise for own-gender faces (Lovén et al., 2011) would lead to an increased tendency to use shortcut strategies such as visual averaging. On this view, the own-gender bias often reported for face recognition (e.g., Cross et al., 1971) and the increased averaging effects found here may both be related to a greater familiarity with own-gender faces. Further work is needed to investigate this issue.

References

- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*, 157–162.
- Barnett-Cowan, M., Dyde, R. T., Thompson, C., & Harris, L. R. (2010). Multisensory determinants of orientation perception: task-specific sex differences. *European Journal of Neuroscience, 31*, 1899–1907.
- Bruce, V., Doyle, T., Dench, N., & Burton, M. (1991). Remembering facial configurations. *Cognition, 38*, 109–144.
- Cabeza, R., Bruce, V., Kato, T., & Oda, M. (1999). The prototype effect in face recognition: extension and limits. *Memory and Cognition, 27*, 139–51.
- Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research, 45*, 891–900.
- Cross, J. F., Cross, J., & Daly, J. (1971). Sex, race, age, and beauty as factors in recognition of faces. *Perception & Psychophysics, 10*, 393–396.
- De Fockert, J. W., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology, 62*, 1716–1722.
- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17*, 751–753.
- Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 718–734.

¹ To what extent are high endorsement rates for morphs in our task indeed indicative of more accurate mean representations? Although our methodology does not include a direct test to confirm that endorsement of matching morphs reflects face averaging, the fact that the endorsement rate for matching morphs is significantly greater than that for nonmatching morphs suggests that a morph is endorsed only when it is the average of the four faces just seen in the preceding set. This effect is most parsimoniously explained by claiming that observers form a summary description of the set of faces, which they use to make their decision as to whether or not a test face may have been present in the set. Converging evidence from other methods testing the nature of the representation of sets of seen faces is needed to further test the claim that average representations are different for own-gender versus other-gender faces.

- Ino, T., Nakai, R., Azuma, T., Kimura, T., & Fukuyama, H. (2010). Gender differences in brain activation during encoding and recognition of male and female faces. *Brain Imaging and Behavior*, *4*, 55–67.
- Jenkins, R., & Burton, A. M. (2008). 100% accuracy in automatic face recognition. *Science*, *319*, 435.
- Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society B*, *366*, 1671–1683.
- Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition: women's faces make a difference. *Brain and Cognition*, *50*, 121–128.
- Lovén, J., Herlitz, A., & Rehnman, J. (2011). Women's own-gender bias in face recognition memory: The role of attention at encoding. *Experimental Psychology*, *58*, 333–340.
- McKelvie, S. J., Standing, L., St. Jean, D., & Law, J. (1993). Gender differences in recognition memory for faces and cars: Evidence for the interest hypothesis. *Bulletin of the Psychonomic Society*, *31*, 447–448.
- Megreya, A. M., Bindemann, M., & Havard, C. (2011). Sex differences in unfamiliar face identification: evidence from matching tasks. *Acta Psychologica*, *137*, 83–89.
- Park, B., & Judd, C. M. (1990). Measures and models of perceived group variability. *Journal of Personality and Social Psychology*, *59*, 173–191.
- Phillips, W. A., Chapman, K. L. S., & Berry, P. D. (2004). Size perception is less context sensitive in males. *Perception*, *33*, 79–86.
- Quattrone, G. A., & Jones, E. E. (1980). The perception of variability within in-groups and out-groups: Implications for the law of small numbers. *Journal of Personality and Social Psychology*, *38*, 141–152.
- Rehnman, J., & Herlitz, A. (2006). Higher face recognition ability in girls: Magnified by own-sex and own-ethnicity bias. *Memory*, *14*, 289–296.
- Rehnman, J., & Herlitz, A. (2007). Women remember more faces than men do. *Acta Psychologica*, *124*, 344–355.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime Reference Guide*. Pittsburgh: Psychology Software Tools Inc.
- Shaw, J. I., & Skolnick, P. (1994). Sex-differences, weapon focus, and eyewitness reliability. *Journal of Social Psychology*, *134*, 413–420.
- Shaw, J. I., & Skolnick, P. (1999). Weapon focus and gender differences in eyewitness accuracy: arousal versus salience. *Journal of Applied Social Psychology*, *29*, 2328–2341.
- Wright, D. B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychologica*, *114*, 101–114.