Research Report

Psychological Science
21(6) 779–785

DOI: 10.1177/0956797610371965
http://pss.sagepub.com

Generalization of Affective Learning About Faces to Perceptually Similar Faces

Sara C. Verosky and Alexander Todorov
Princeton University

Abstract

Different individuals have different (and different-looking) significant others, friends, and foes. The objective of this study was to investigate whether these social face environments can shape individual face preferences. First, participants learned to associate faces with positive, neutral, or negative behaviors. Then, they evaluated morphs combining novel faces with the learned faces. The morphs (65% and 80% novel faces) were within the categorical boundary of the novel faces: They were perceived as those faces in a preliminary study. Moreover, a second preliminary study showed that following the learning, the morphs' categorization as similar to the learned faces was indistinguishable from the categorization of actual novel faces. Nevertheless, in the main experiment, participants evaluated morphs of "positive" faces more positively than morphs of "negative" faces. This learning generalization effect increased as a function of the similarity of the novel faces to the learned faces. The findings suggest that general learning mechanisms based on similarity can account for idiosyncratic face preferences.

Keywords

face perception, social learning, person perception

Received 8/10/09; Revision accepted 11/13/09

Although there is evidence that people agree when judging other people from facial appearance (Hassin & Trope, 2000; Todorov, Said, Engell, & Oosterhof, 2008; Zebrowitz & Montepare, 2008), there is also evidence that these judgments reflect not only properties of the face, but also properties of the judge (Engell, Haxby, & Todorov, 2007; Hönekopp, 2006). It is unlikely that the latter idiosyncratic contributions to judgments simply reflect random noise across judges. For example, people have more positive evaluations of faces that resemble themselves than of faces that have no such resemblance (Bailenson, Iyengar, Yee, & Collins, 2009; DeBruine, 2002, 2005). Thus, it seems that part of beauty really is in the "eye of the beholder."

We argue that evaluation of novel faces is partially based on their similarity to familiar faces, a hypothesis consistent with the familiar-face overgeneralization hypothesis (Zebrowitz, 1996; Zebrowitz & Collins, 1997). Different individuals have different social interactions, and everyday learning from such interactions could influence the evaluation of novel faces. For example, in an early study, Lewicki (1985) showed that a short pleasant or unpleasant interaction with an experimenter affected participants' choices of a new experimenter. Participants chose the person who resembled the experimenter (both had short hair and eyeglasses) when the interaction was pleasant, but chose the other person (who had long hair and no glasses) when the interaction was unpleasant.

Our goal in the study reported here was to test whether rapidly learned affective associations with faces are generalized to novel faces that resemble the familiar faces. Finding that evaluation of novel faces is influenced by their similarity to familiar faces would suggest that "unique" individual face preferences might result, in part, from a common underlying process of learning generalization.

To design the experiment, we capitalized on two phenomena: categorical perception of faces (Beale & Keil, 1995; Levin & Beale, 2000) and findings that social judgments of a face are easily changed when people are provided with information about the person (Bliss-Moreau, Barrett, & Wright, 2008; Goren & Todorov, 2009; Todorov & Olson, 2008; Todorov & Uleman, 2003). Categorical perception of facial identity is indicated by a sharp boundary in perception of morphs of faces that vary continuously (Levin & Beale, 2000). Differences between morphs within the identity boundary of a face are attenuated, whereas differences between morphs on different sides of the category boundary are accentuated.

Corresponding Author:
Alexander Todorov, Department of Psychology, Princeton University, Princeton, NJ 08540
E-mail: atodorov@princeton.edu

In the main experiment, participants learned to associate faces with positive, neutral, or negative behavioral information. Then, they evaluated morphs that combined these faces with novel faces and were within the categorical boundaries of the novel faces. Consistent with prior findings (Jacques & Rossion, 2006; Levin & Beale, 2000), a preliminary study found that morphs containing more than 65% of a given face were identified as that face (Fig. 1). Thus, in the main experiment, participants evaluated novel faces that were morphed with 20% or 35% of the learned faces. In an additional preliminary study, we further showed that participants treated the morphs like completely novel faces. Specifically, after participants learned the associations of faces and behaviors, their categorization of the morphs as similar to the learned faces

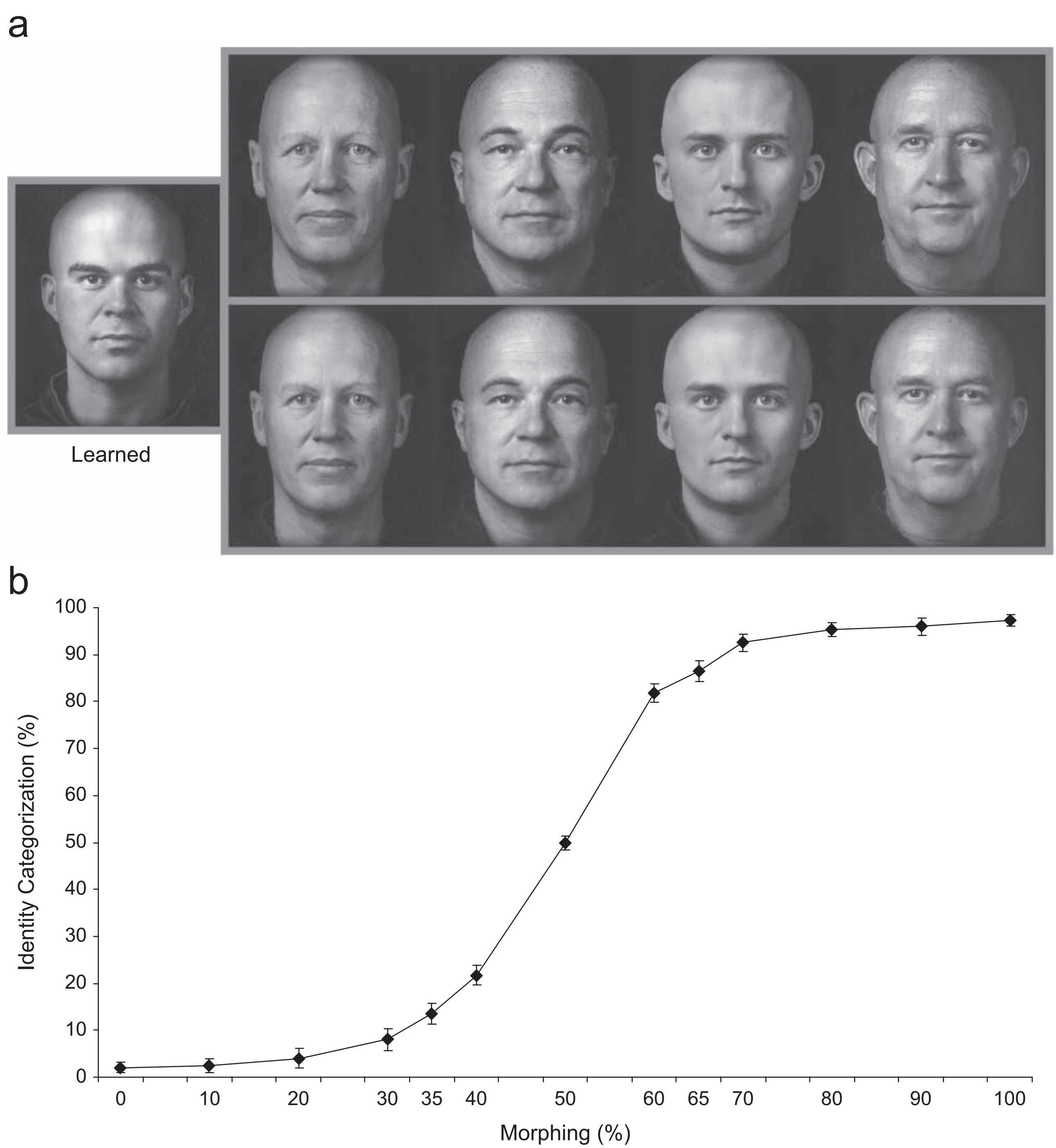

Fig. 1. The category boundary between faces. Examples of stimuli used in the experiment are presented in (a). The top row shows morphs containing 20% of the learned face in the box on the left, and the bottom row shows morphs containing 35% of the learned face. The graph (b) shows the percentage of the time that participants (N = 16) in Preliminary Study 1 identified morphs as looking like the original face as a function of the percentage of the original face present. Error bars represent standard errors of the mean.

was indistinguishable from their categorization of actual novel faces as similar to the learned faces.

Despite the low similarity of the morphs to the learned faces, we expected that those faces' behavioral associations would influence the evaluation of the morphs. Specifically, we expected that novel faces morphed with "positive" learned faces would be evaluated more positively than novel faces morphed with "negative" learned faces. We also expected that this learning generalization effect would increase as a function of similarity. That is, we predicted that the effect would be stronger for 35% than for 20% morphs.

Method

Participants

Thirty-one undergraduate students participated in the preliminary studies, and another 57 participated in the main study for partial course credit.

Stimuli

We used 54 photographs of men with neutral facial expressions from a set of black-and-white photographs of bald males (Kayser, 1997). For Preliminary Study 2 and the main experiment, we selected 9 photographs of young to middle-aged men to be used as the learned faces. These 9 faces were divided into three groups, and participants learned to associate each group with positive, neutral, or negative behaviors. The behaviors were taken from a database of 400 behaviors that includes trait ratings of those behaviors; the behaviors were selected on the basis of their goodness ratings (Fuhrman, Bodenhausen, & Lichtenstein, 1989). The behavior associated with each face group was counterbalanced across participants.

Procedure

Preliminary Study 1. To determine the location of the identity boundary between pairs of faces, we created a series of morphs. Faces were morphed in increments of 10%, using code from Steyvers (1999). In addition, morphs containing 35% and 65% of the base face were included because the greatest change in perceived identity was expected to occur close to these positions. Participants (N = 16) were asked to categorize each morph as looking like one or the other endpoint face (see the Supplemental Material available online).

Preliminary Study 2. This study began with a learning phase, in which participants (N = 15) were told that their task was to form person impressions. They were presented with nine faces paired with behavioral descriptions (e.g., "he stole money and jewelry from the relatives he was living with") and asked to imagine the people pictured performing the behaviors—a procedure adapted from previous studies (Bliss-Moreau et al., 2008; Todorov & Uleman, 2003). Each face was paired with three different behaviors of the same valence: positive, neutral, or negative. The faces were blocked so that participants saw each of the nine faces paired with a first behavior, then each of the nine faces paired with a second behavior, and finally each of the nine faces paired with a third behavior. Each time, the faces were shown in a different random order. The presentation of the face-behavior pairs was self-paced, and the intertrial interval was 1,000 ms.

After seeing each face paired with three behaviors, participants saw the faces alone and were asked to indicate whether each one was previously presented with positive, neutral, or negative behaviors. Participants received feedback about their accuracy. If participants gave an incorrect response, they saw each of the nine faces paired with another behavior of the same valence and then completed another test round. Participants continued this procedure until they reached 100% accuracy or completed eight test rounds. Each face was paired with at most five different behaviors before the behaviors were repeated.

After the learning phase, participants were told that they would see learned faces, faces similar to the learned faces, and novel faces. They were asked to categorize each face as learned, similar to learned, or new. Each participant saw 9 learned faces, 36 morphed faces, and 9 novel faces in a random order. On the basis of the results from Preliminary Study 1 (Fig. 1b), we decided to use 20% and 35% morphs of the learned faces as the morphs in this study. Each of the 9 learned faces was morphed with 4 of 36 novel faces at two different levels of morphing (20% and 35%). Half of the novel faces were shown at one morphing level and half at the other; the morphing level used for a given face was counterbalanced across participants.

Main experiment. The main experiment began with a learning phase, which followed the same procedure as in Preliminary Study 2. In the following evaluation phase, participants were told that we were interested in first impressions and that they would evaluate both learned and novel faces. Unlike participants in Preliminary Study 2, participants in the main experiment were not told that they would see faces that were similar to the learned faces. Prior work has demonstrated that trustworthiness judgments provide a good approximation of valence evaluation of faces (Oosterhof & Todorov, 2008); therefore, we asked participants to rate the faces on trustworthiness. The morphed faces were the same as those used in the evaluation phase of Preliminary Study 2 (see the Supplemental Material online).

Each trial started with a 500-ms fixation cross before a face was displayed. The face remained on the screen until the participant rated its trustworthiness on a scale from 1 (*not at all trustworthy*) to 9 (*extremely trustworthy*), using the number keys on a computer keyboard. Each participant rated all faces twice, and each time the faces were shown in a different random order.

Results

Preliminary studies

Results from Preliminary Study 1 showed that at the 20% morphing level, the similarity between the morphs and the original face was difficult to detect (Fig. 1a), and the morphs were almost always categorized as looking like the other endpoint face (Fig. 1b). At the 35% morphing level, the morphs were categorized as looking like the latter face more than 85% of the time.

The second preliminary study further showed that the morphs were treated as novel faces (Fig. 2). On the majority of trials, participants categorized the morphs as novel faces (M = 74%, SD = 16%); they almost never categorized the morphs as learned faces (M = 2%, SD = 3%). These percentages were not significantly different from the respective percentages for actual novel faces, ts < 1. Although participants sometimes categorized the morphs as similar to the learned faces (M = 24%, SD = 16%), this percentage was not different from the percentage of novel faces categorized as similar to the learned

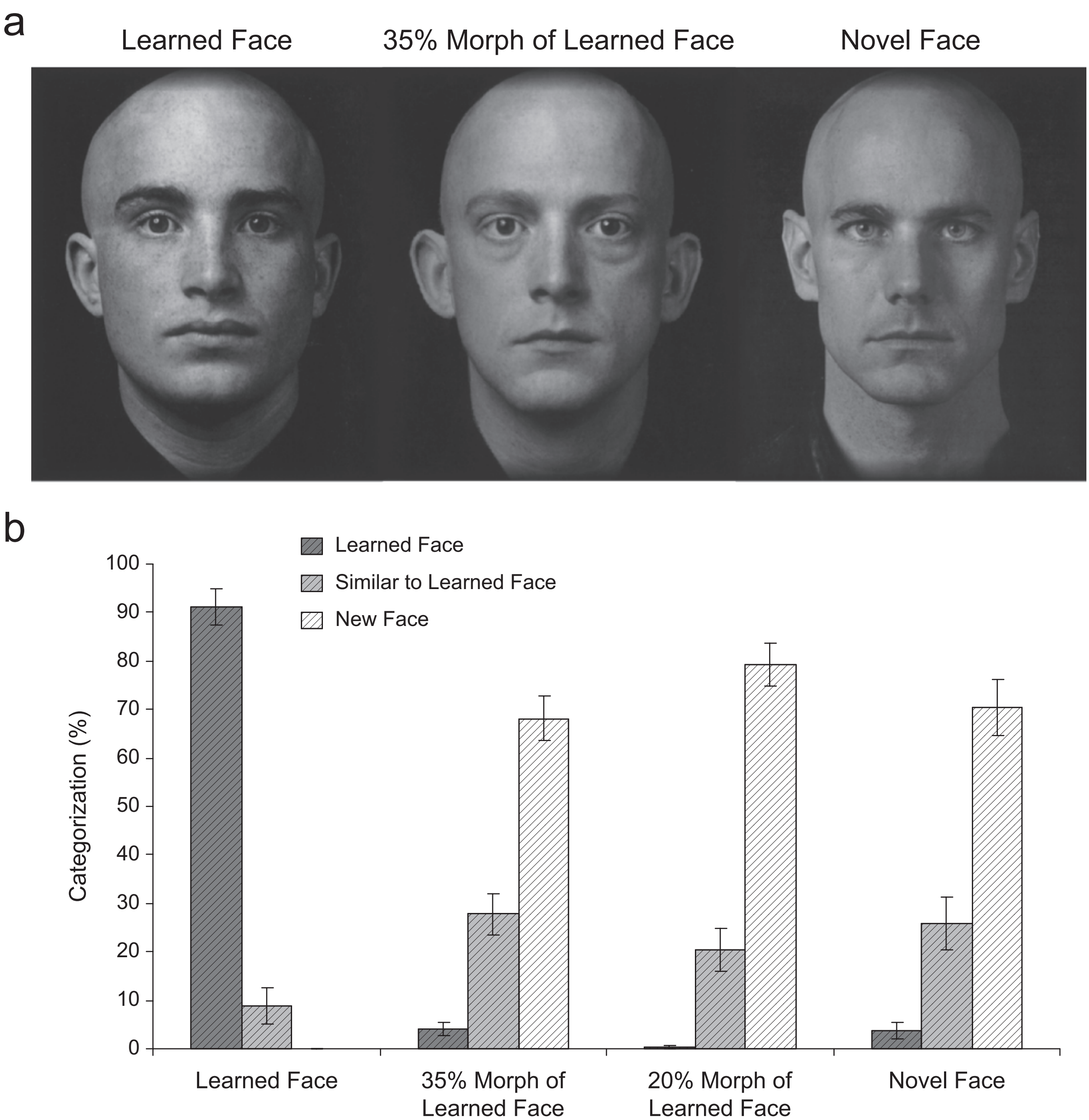

Fig. 2. Examples of stimuli (a) and results (b) from Preliminary Study 2. The examples show a learned face, a 35% morph of this face, and a completely novel face. The graph shows the percentage of the time that participants categorized faces as learned, similar to learned, or new as a function of type of face (N = 15). Error bars represent standard errors of the mean.

faces, $t < 1$ (see the Supplemental Material online for additional analyses).

Main experiment

We analyzed the results of the main experiment using 3 (valence of learning: positive vs. neutral vs. negative) × 2 (morphing: 35% vs. 20% of the learned face) repeated measures analysis of variance.

Learning. In the learning phase, slightly more than 50% of participants correctly identified the face-behavior associations after two test rounds, 75% were able to do so after four rounds, and 100% were able to do so after eight rounds. Analysis of the judgments of the learned faces revealed a large effect of the valence of the behavior, $F(2, 112) = 93.57$, $p < .001$, $\eta_p^2 = .63$. Participants evaluated faces that were previously associated with negative behaviors ($M = 3.02$, $SD = 1.35$) more negatively than faces that were associated with neutral behaviors ($M = 5.50$, $SD = 1.15$), $t(56) = 10.90$, $p < .001$. In turn, they rated the latter more negatively than faces that were associated with positive behaviors ($M = 6.56$, $SD = 1.59$), $t(56) = 4.54$, $p < .001$. Only 3 of the 57 participants did not rate faces associated with positive behaviors more positively than faces associated with negative behaviors. The remaining analyses were conducted on the data from the 54 participants who showed evidence of learning.[1] It should be noted that this is a liberal criterion of learning. By a more strict criterion, the judgments should reflect strict ranking of the associations as follows: positive > neutral > negative. Only 35 participants exhibited this pattern.

Learning generalization. As expected, participants rated novel faces morphed with positive learned faces ($M = 5.25$, $SD = 0.94$) more positively than novel faces morphed with neutral ($M = 5.20$, $SD = 0.87$) or negative ($M = 4.99$, $SD = 0.90$) faces, $F(2, 106) = 3.09$, $p < .05$, $\eta_p^2 = .06$. There was also a significant main effect of morphing; 35% morphs were evaluated more positively than 20% morphs, $F(1, 53) = 11.14$, $p < .002$, $\eta_p^2 = .17$. These two main effects were qualified by a significant interaction, $F(2, 106) = 3.81$, $p < .03$, $\eta_p^2 = .07$, which reflected the fact that the learning generalization effect was more pronounced for the 35% than for the 25% morphs.

In the 20% morph condition, participants did evaluate morphs of positive faces more positively than morphs of negative faces (Fig. 3a), but the effect was not significant, Fs < 1 for the overall effect and the linear contrast. When the analysis was limited to the 35 participants who showed the expected relative ranking of learned faces according to their valence, the linear contrast reached marginal significance (Fig. 3b), $F(1, 34) = 3.46$, $p < .071$, $\eta_p^2 = .09$. For the 35% morphs, the effect of behavior valence was significant (Fig. 3c), $F(2, 106) = 5.93$, $p < .004$, $\eta_p^2 = .10$, for the overall effect and $F(1, 53) = 9.59$, $p < .003$, $\eta_p^2 = .15$, for the linear contrast. The linear effect was stronger for participants who showed the expected relative ranking of learned faces (Fig. 3d), $F(1, 34) = 10.48$, $p < .003$, $\eta_p^2 = .24$.

Discussion

Our findings provide a laboratory demonstration of a phenomenon that we suspect occurs in everyday life. The evaluation of novel faces is affected by their perceptual similarity to familiar faces. This similarity becomes a vehicle for transfer of affective associations from the familiar to the novel faces. In the main experiment, participants first learned to associate faces with behaviors that varied in valence and then evaluated a second set of faces, which had been subtly manipulated to resemble the learned faces. Even though the learning that took place in this experiment pales in comparison with the type of learning that occurs in the real world, it affected participants' impressions of the novel faces. Faces that resembled positive faces were evaluated more positively than faces that resembled negative faces. This learning generalization effect increased as a function of the similarity of the novel to the learned faces. The finding that participants did not perceive the morphed faces differently from completely novel faces suggests that conscious recognition of similarity may not be necessary for learning generalization. However, more work is needed to explore this possibility.

Our findings provide additional experimental support for the familiar-face overgeneralization hypothesis (Zebrowitz, 1996). Previous tests of this hypothesis have focused on familiarity at a group level, demonstrating that greater familiarity with in-group (within-race) faces relative to out-group faces can partially explain in-group face preferences (Zebrowitz, Bronstad, & Lee, 2007; Zebrowitz, Wieneke, & White, 2008). The current study focused on the familiarity of specific individuals. In this respect, our findings are related to Andersen and her colleagues' findings that a perceiver's impressions of unfamiliar others are influenced by their resemblance to the perceiver's significant others (Andersen & Cole, 1990; Chen & Andersen, 1999). Most likely, one of the proximal cues for this process is facial similarity (cf. Kraus & Chen, 2010). However, it should be noted that our account is broader than Andersen's account. In the current study, evaluation of novel faces was affected by their similarity to faces that had been learned quickly and had no personal significance for the participants.

Finally, our account is related to work suggesting that learning of incidental associations between facial features and personality characteristics could contribute to individual differences in face preferences. For example, Hill, Lewicki, Czyzewska, and Schuller (1990) created a covariation between the length of a set of faces and a behavioral trait and found that this association influenced participants' ratings of a second set of faces. In a related study, Jones, DeBruine, Little, and Feinberg (2007) found that participants preferred composite faces made up of individual faces that were previously paired with neutral sounds over composites made up of

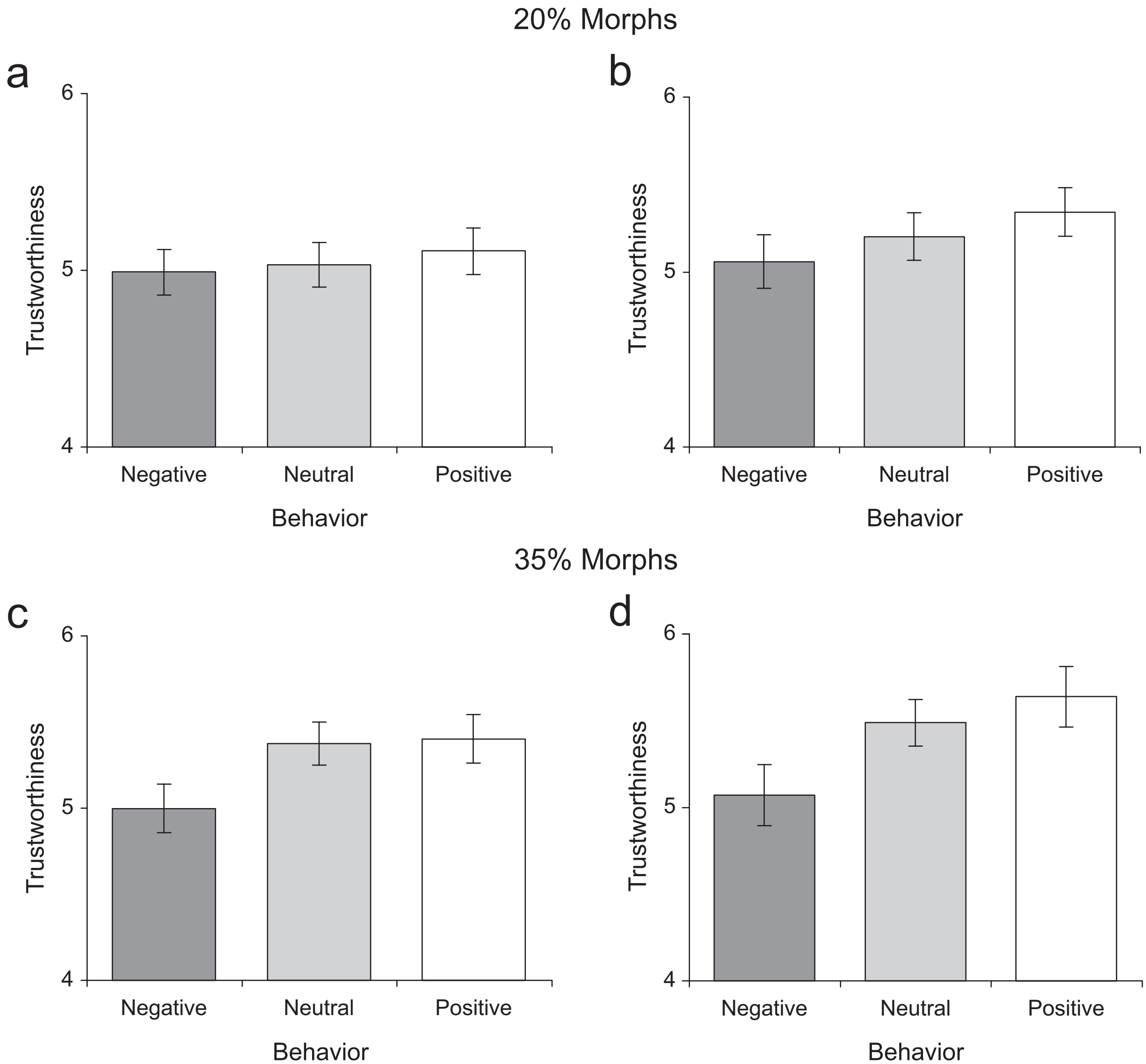

Fig. 3. Behavioral results demonstrating learning generalization: evaluation of morphs as a function of the valence of the behavior associated with the learned face. The graphs in the top row show the mean trustworthiness ratings of the morphs containing 20% of the learned faces among (a) the 54 participants who showed evidence of learning and (b) the 35 participants who showed the expected relative ranking of learned faces according to their valence. The graphs in the bottom row show the mean trustworthiness ratings of the morphs containing 35% of the learned faces among (c) the 54 participants who showed evidence of learning and (d) the 35 participants who showed the expected relative ranking of learned faces according to their valence. Error bars represent standard errors of the mean.

faces previously paired with negative sounds. The extent to which the processes underlying the generalization of these kinds of associations differ from the processes underlying the generalization of associations based on person-specific attributional learning (e.g., Todorov & Uleman, 2003) remains an open question.

Most prior work on face evaluation has focused on what different people's judgments have in common (Eagly, Ashmore, Makhijani, & Longo, 1991; Todorov et al., 2008). However, there is substantial individual variation in these judgments (Hönekopp, 2006). The current study investigated a potential source of this variation. Different individuals have different (and different-looking) significant others, friends, and foes. This social face environment would most likely shape preferences for novel faces. In sum, our findings suggest that general learning mechanisms based on similarity can account for idiosyncratic face preferences.

Acknowledgments

We thank Jenny Porter for her help with the data collection.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was supported by National Science Foundation Grant BCS-0823749.

Supplemental Material

Additional supporting information may be found at http://pss.sagepub.com/content/by/supplemental-data

Note

1. Including the 3 excluded participants did not affect the results.

References

Andersen, S.M., & Cole, S.W. (1990). "Do I know you?": The role of significant others in general social perception. *Journal of Personality and Social Psychology*, *59*, 384–399.

Bailenson, J.N., Iyengar, S., Yee, N., & Collins, N.A. (2009). Facial similarity between voters and candidates causes influence. *Public Opinion Quarterly*, *72*, 935–961.

Beale, J.M., & Keil, F.C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*, 217–239.

Bliss-Moreau, E., Barrett, L.F., & Wright, C.I. (2008). Individual differences in learning the affective value of others under minimal conditions. *Emotion*, *8*, 479–493.

Chen, S., & Andersen, S.M. (1999). Relationships from the past in the present: Significant-other representations and transference in interpersonal life. In M.P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 31, pp. 123–190). San Diego, CA: Academic Press.

DeBruine, L.M. (2002). Facial resemblance enhances trust. *Proceedings of the Royal Society B: Biological Sciences*, *269*, 1307–1312.

DeBruine, L.M. (2005). Trustworthy but not lust-worthy: Context-specific effects of facial resemblance. *Proceedings of the Royal Society B: Biological Sciences*, *272*, 919–922.

Eagly, A.H., Ashmore, R.D., Makhijani, M.G., & Longo, L.C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*, 109–128.

Engell, A.D., Haxby, J.V., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience*, *19*, 1508–1519.

Fuhrman, R.W., Bodenhausen, G.V., & Lichtenstein, M. (1989). On the trait implications of social behaviors: Kindness, intelligence, goodness, and normality ratings for 400 behavior statements. *Behavior Research Methods, Instruments, & Computers*, *21*, 587–597.

Goren, A., & Todorov, A. (2009). Two faces are better than one: Eliminating false trait associations with faces. *Social Cognition*, *27*, 222–248.

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, *78*, 837–852.

Hill, T., Lewicki, P., Czyzewska, M., & Schuller, G. (1990). The role of learned inferential encoding rules in the perception of faces: Effects of nonconscious self-perpetuation of a bias. *Journal of Experimental Social Psychology*, *26*, 350–371.

Hönekopp, J. (2006). Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 199–209.

Jacques, C., & Rossion, B. (2006). The speed of individual face categorization. *Psychological Science*, *17*, 485–492.

Jones, B.C., DeBruine, L.M., Little, A.C., & Feinberg, D.R. (2007). The valence of experience with faces influences generalized preferences. *Journal of Evolutionary Psychology*, *5*, 119–129.

Kayser, A. (1997). *Heads*. New York: Abbeville Press.

Kraus, M.W., & Chen, S. (2010). Facial-feature resemblance elicits the transference effect. *Psychological Science*, *21*, 518–522.

Levin, D.L., & Beale, J.M. (2000). Categorical perception occurs in newly learned faces, other-race faces and inverted faces. *Perception & Psychophysics*, *62*, 386–401.

Lewicki, P. (1985). Nonconscious biasing effects of single instances on subsequent judgments. *Journal of Personality and Social Psychology*, *48*, 563–574.

Oosterhof, N.N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, USA*, *105*, 11087–11092.

Steyvers, M. (1999). Morphing techniques for generating and manipulating face images. *Behavior Research Methods, Instruments, & Computers*, *31*, 359–369.

Todorov, A., & Olson, I.R. (2008). Robust learning of affective trait associations with faces when the hippocampus is damaged, but not when the amygdala and temporal pole are damaged. *Social Cognitive and Affective Neuroscience*, *3*, 195–203.

Todorov, A., Said, C.P., Engell, A.D., & Oosterhof, N.N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*, 455–460.

Todorov, A., & Uleman, J.S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, *39*, 549–562.

Zebrowitz, L.A. (1996). Physical appearance as a basis of stereotyping. In N. Macrae, M. Hewstone, & C. Stangor (Eds.). *Foundations of stereotypes and stereotyping* (pp. 79–120). New York: Guilford Press.

Zebrowitz, L.A., Bronstad, P.M., & Lee, H.K. (2007). The contribution of face familiarity to ingroup favoritism and stereotyping. *Social Cognition*, *25*, 306–338.

Zebrowitz, L.A., & Collins, M.A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review*, *1*, 204–223.

Zebrowitz, L.A., & Montepare, J.M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*, 1497–1517.

Zebrowitz, L.A., Wieneke, K., & White, B. (2008). Mere exposure and racial prejudice: Exposure to other-race faces increases liking for strangers of that race. *Social Cognition*, *26*, 259–275.