Journal of Experimental Social Psychology 49 (2013) 661–669

Contents lists available at SciVerse ScienceDirect

Journal of Experimental Social Psychology

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

When physical similarity matters: Mechanisms underlying affective learning generalization to the evaluation of novel faces

Sara C. Verosky *, Alexander Todorov

Department of Psychology, Princeton University, Princeton, NJ 08540, USA

HIGHLIGHTS

- ► Participants learned to associate faces with behaviors and judged morphs of the faces.
- ► Learning generalized to morphs despite presence of relevant behavioral information
- ► Learning generalized while participants were under cognitive load
- ► Learning generalized despite instructions to disregard physical similarity
- ► Learning generalization is a powerful and relatively automatic process

ARTICLE INFO

Article history:
Received 7 August 2012
Revised 7 February 2013
Available online 16 February 2013

Keywords:
Face learning
Perceptual similarity
Overgeneralization
Trait impressions
Automaticity

ABSTRACT

In two experiments, participants first learned to associate faces with negative, neutral, or positive behaviors and then evaluated morphs of these faces with novel faces. Across both experiments, participants evaluated new (morph) faces that were similar to familiar negative faces more negatively than new (morph) faces that were similar to familiar positive faces. This learning generalization effect was present when participants' judgments of the new (morph) faces were a) based not only on facial appearance but also on relevant behavioral information (Experiment 1); b) made under cognitive load (Experiment 2); and c) made under instructions not to use similarity information (Experiment 2). The findings of the experiments suggest that learning generalization based on facial physical similarity is a powerful and relatively automatic process, which likely influences face evaluation across a range of circumstances.

Introduction

Similarity plays a fundamental role in perception, learning, and judgment. In fact, similarity is so central to our thinking that it is difficult to define. As Quine (1969) noted, "The dubiousness of this notion is itself a remarkable fact. For surely there is nothing more basic to thought and language than our sense of similarity; our sorting of things into kinds" (pg. 116). Within the domain of face perception, physical similarity is important for both face recognition and face categorization, and serves as a basic organizing principle of models of face representation (O'Toole, 2011; Rhodes & Jeffery, 2006; Rhodes & Leopold, 2011; Valentine, 1991).

Physical similarity also plays an important role in face evaluation: even when a face does not belong to a category, physical similarity to that category can act as a cue influencing people's impressions of that person. In line with this idea, people's impressions of others have been found to be sensitive to similarity along a number of dimensions, including health, age, emotion, and gender (Oosterhof & Todorov, 2008; Todorov, Said, Engell, & Oosterhof, 2008; Zebrowitz, 1996; Zebrowitz & Montepare, 2008). For example, using connectionist modeling, Zebrowitz, Fellous, Mignault, and Andreoletti (2003) found that judgments of attractiveness vary with the physical similarity of normal faces to faces with genetic anomalies, while judgments of babyfaceness vary with similarity of adult faces to the faces of babies. Similarly, both computational modeling and behavioral studies suggest that trait judgments of emotionally neutral faces result from similarity of those faces to faces expressing emotion (Montepare & Dobish, 2003; Oosterhof & Todorov, 2008, 2009; Said, Sebe, & Todorov, 2009; Zebrowitz, Kikuchi, & Fellous, 2010). To give an example, when Oosterhof and Todorov (2008) created a model representing people's trustworthiness judgments of emotionally neutral faces, they found that faces at extreme ends of the trustworthiness spectrum looked like they were expressing anger and happiness, respectively.

Although the studies mentioned above investigated how different types of physical similarity influence initial evaluation of faces, there is also evidence that these cues continue to color impressions even in more naturalistic interactions. Looking at cases in small claims court, Zebrowitz and McDonald (1991) found that babyfaceness and attractiveness influenced legal outcomes, despite the fact that judges are supposed to be making decisions based solely on the evidence

* Corresponding author.
E-mail address: sverosky@princeton.edu (S.C. Verosky).

0022-1031/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jesp.2013.02.004

presented to them. More generally, a series of meta-analyses found that attractive children and adults are judged and treated more positively than unattractive children and adults in a range of settings, even by those who know them (Langlois et al., 2000). Thus, appearance not only matters for initial impressions, but also continues to exert an influence even as interactions unfold.

The current work investigates a type of similarity rooted in individual experience: similarity to familiar others. Physical similarity to familiar others has previously been shown to influence impressions based on facial appearance (DeBruine, 2002, 2005; Günaydin, Zayas, Selcuk, & Hazan, 2012; Hill, Lewicki, Czyzewska, & Schuller, 1990; Kraus & Chen, 2010; Lewicki, 1985; Tanner & Maeng, 2012; Verosky & Todorov, 2010; Zebrowitz, 1996; Zebrowitz, Bronstad, & Lee, 2007; Zebrowitz, Wieneke, & White, 2008). For instance, people who believed that they were going to interact with someone whose facial features resembled their significant other were likely to infer that this interaction partner possessed characteristics of their significant other (Kraus & Chen, 2010). While this finding illustrates that facial resemblance can trigger transference of significant others' characteristics to strangers (Andersen & Cole, 1990), other work indicates that the processes by which physical similarity influences person evaluation need not be limited to significant others (Hill et al., 1990; Lewicki, 1985).

In Verosky and Todorov (2010), participants learned to associate faces with negative, neutral, and positive behaviors. Then, participants evaluated the trustworthiness of a series of morphs of novel faces with the familiar faces. Participants generalized their impressions of the familiar faces to the similar-looking novel faces, such that novel faces similar to faces associated with negative behaviors were evaluated more negatively than novel faces similar to faces associated with positive behaviors.

While Verosky and Todorov (2010) demonstrated that it is possible for social face environments to shape face preferences, it did not speak to the mechanisms underlying the effect. In the current work, we used the paradigm developed in Verosky and Todorov (2010) to investigate the possibility that use of physical similarity in the evaluation of faces is relatively automatic. Given the importance of physical similarity for face processing, it is not difficult to imagine that information about similarity to familiar others is readily available to people as they evaluate novel faces.

In the first experiment, we were interested in whether physical similarity to known others would continue to influence impressions in the presence of relevant behavioral information. If similarity to known others is treated in the same way as other appearance-related information, we would expect that it would continue to influence impressions. If however, use of physical similarity is more deliberate, we might expect that the effect would be dampened or disappear in the presence of information diagnostic for the impression.

In the second experiment, we used two experimental manipulations to explore participants' control over their use of similarity. First, in order to see whether participants could change their use of similarity information, we instructed them to either include similarity to familiar others in their judgments or to disregard it and we compared learning generalization in those conditions to learning generalization in a baseline condition without instructions. Second, we introduced a cognitive load manipulation: to the extent that use of similarity information is efficient, load should not interfere with learning generalization.

Experiment 1

As a starting point, Experiment 1 examined whether people would continue to use physical similarity in their judgments even when there was other information available. Although appearance is often one of the first cues observers have about another person, information about their actions becomes increasingly available over time. Past research has shown that these actions have large effects on observers' evaluations of the person (Bliss-Moreau, Barrett, & Wright, 2008; Carlston & Skowronski, 1994; Carlston, Skowronski, & Sparks, 1995; Crawford, Skowronski, Stiff, & Scherer, 2007; Skowronski, Carlston, Mae, & Crawford, 1998; Todorov & Uleman, 2002, 2003, 2004). This raises the following question: Is physical similarity to a known other simply something people rely on when there is no other information available or does it continue to color impressions as people learn more information about the person in question?

Physical similarity does not provide evidence as to the true underlying similarity between individuals, but a person's own actions can give valuable clues as to their character. Therefore, to the extent that use of physical similarity in evaluation is strategic, we might expect that people will cease to rely on this information as other information becomes available. Along these lines, Lewicki (1985) examined use of physical resemblance as a case study for what people do when they need to make a decision, but do not have sufficient information to do so. He found that when participants had a pleasant versus unpleasant interaction with an initial experimenter this affected whether they chose to interact with a second experimenter who resembled the first (both had short hair and eye glasses). While the assumption in this study was that the physical resemblance would only matter insofar as there was no other information available, this was never explicitly tested.

On the other hand, to the extent that physical similarity is treated like other sources of appearance related information, we might expect that it would continue to matter even in the presence of other behavioral information. People tend to show agreement in their judgments based on facial appearance and such appearance related information is integrated with behavioral information (Baron, Gobbini, Engell, & Todorov, 2011; Rudoy & Paller, 2009; Todorov & Olson, 2008). For example, Todorov and Olson (2008) presented trustworthy- and untrustworthy-looking faces with negative and positive behaviors and found that both the appearance of a face and the valence of behavior influenced evaluation.

The first study was designed to investigate whether learning generalization would persist even in the presence of relevant behavioral information. Participants first saw one face paired with a negative vignette, one with a neutral vignette, and one with a positive vignette (see Appendix A). Then, they were asked to evaluate a series of morphs of novel faces with the familiar faces. These new (morph) faces were presented alone, replicating Verosky and Todorov (2010), or with concurrent negative, neutral, or positive behaviors. Pairing faces with brief behavioral descriptions is a powerful manipulation, which has previously been found to change impressions of others (Bliss-Moreau et al., 2008; Todorov & Uleman, 2002, 2003, 2004). For instance, viewing faces paired with trait-implying behaviors for as little as 2 s leads to forming of corresponding trait associations with the faces (Todorov & Uleman, 2003). Given the strength of this manipulation, we were interested in whether physical similarity to the familiar faces would continue to influence evaluation of the new (morph) faces even when those faces themselves were presented with behavioral information. If participants take both behavior and physical similarity into account in their evaluation of the new (morph) faces, we should see an effect of concurrent behavioral information on evaluation of the new (morph) faces, as well as an effect of learning generalization.

Method

Participants

Seventy-two participants were recruited at a local shopping mall and completed the task at a booth in the mall for payment. Data from one participant who entered the same response for every face were excluded from further analysis.

Stimuli

Stimuli consisted of 39 photographs of men with neutral facial expressions taken from a set of 62 digital black-and-white photographs of bald males (Goldstone & Steyvers, 2001; http://psiexp.ss.uci.edu/research/software.htm). We used bald males because this allowed us to avoid concerns about the realism of morphed hair. These images were originally selected from a book of photographs (Kayser, 1997). We chose three photographs of younger to middle-aged men to use as the familiar faces (Fig. 1). One face was paired with a negative biographical vignette, one with a neutral vignette, and one with a positive vignette (see Appendix A). The pairing of faces and vignettes was counterbalanced across participants.

Each of the three familiar faces was morphed with 12 of the remaining faces to create 36 new (morph) faces total (see Fig. 1 for example stimuli). The faces were morphed to a level of 50%, using code from Steyvers (1999). In Verosky and Todorov (2010), we used levels of morphing of 20% and 35%, with higher morphing leading to a stronger effect of learning generalization. Here, we chose to use a still higher level of morphing to better examine the circumstances under which learning generalization occurred or did not occur. While this is a higher level of morphing than we have used in the past, it is comparable to the levels used in other studies. For example, DeBruine (2002, 2005) morphed the self-face with unknown faces at levels of 40% and 50% (including both color and shape information in the morphs, and shape alone, respectively) and reported that participants did not detect this manipulation.

Three of the morphs of each familiar face were paired with negative behaviors, three with neutral behaviors, three with positive behaviors, and three were presented without behaviors. The behaviors were selected from a database of 400 behaviors based on goodness ratings (Fuhrman, Bodenhausen, & Lichtenstein, 1989). The pairing of faces and behaviors was counterbalanced across participants.

Procedures

Learning. Prior to seeing the faces paired with vignettes, participants were asked to evaluate the trustworthiness of the faces based on appearance alone. Participants were told that we were interested in first impressions and that there were no right or wrong answers. Each trial began with a 250-millisecond fixation cross. Then a face was presented and remained on the screen until participants responded using the number keys 1 (not at all trustworthy) to 9 (extremely trustworthy). After evaluating the faces, participants saw the faces paired with vignettes. Each face-vignette pair remained on the screen until participants pressed the spacebar. After seeing all three pairs, the three faces appeared side by side and participants were asked to indicate which face belonged to the "good guy" and which belonged to the "bad guy". Participants responded using the number key corresponding to the position of the faces. After each response, participants saw the word "correct" or "incorrect". If a participant gave an incorrect response, they saw the faces and biographies again before completing another test round. The learning phase ended when participants answered correctly.

Face evaluation. After completing the learning phase of the experiment, participants were asked to evaluate the trustworthiness of a series of faces. Participants were told that some of the faces would be presented with behaviors and that some would be presented alone. For the faces with behaviors, participants were instructed to form impressions of the people by imagining them actually performing the behaviors. For the faces presented alone, participants were instructed to rely on their gut feeling.

Each trial began with a 250-millisecond fixation cross. Then a face appeared and remained on the screen until participants responded using the number keys 1 (not at all trustworthy) to 9 (extremely trustworthy). The order of trials was randomized for each participant. After evaluating the 36 new (morph) faces, participants were told that they would see the faces from the beginning of the experiment. Participants evaluated these faces following the same procedure. Although we also collected ratings of the familiar faces prior to the presentation of biographical information, learning in this and the subsequent experiment was assessed based on the ratings collected at the end of the experiment. This was done because behavioral information had a strong effect on face evaluation and learning assessed

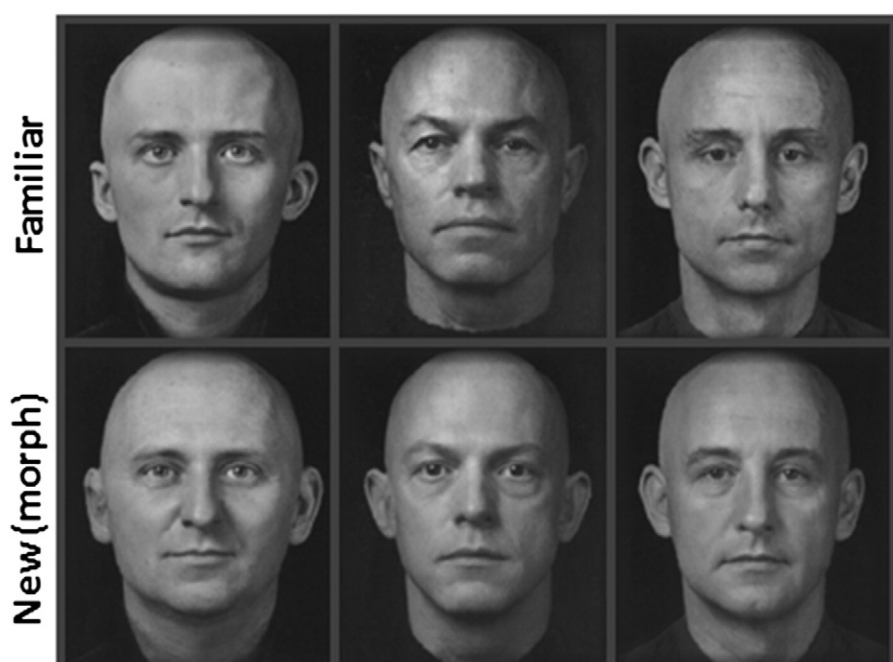

Fig. 1. New (morph) faces. The three familiar faces (top row) and examples of morphs containing 50% of those faces (bottom row). Morphs are shown directly below the faces they contain.

by comparing pre- and post-ratings was largely redundant with learning assessed by comparing post-ratings alone[1].

The overall design was a 3 (valence of learning: negative, neutral, or positive) × 4 (concurrent information: no information, negative, neutral, or positive) repeated measures analysis of variance (ANOVA).

Results

Manipulation check

Sixty-nine percent of participants were able to identify the "good guy" and the "bad guy" after the first test round, 85% of participants were able to do so after the second test round, and all participants were able to do after the sixth test round. The biographical information influenced participants' evaluation of the familiar faces, $F(2, 140)=28.09$, $p<.001$, $\eta_p^2=.29$, with participants evaluating the face paired with the negative biography ($M=3.63$, $SD=2.36$) significantly more negatively than the face paired with the neutral ($M=5.68$, $SD=1.60$; $t(70)=6.53$, $p<.001$) or positive biography ($M=6.25$, $SD=2.18$; $t(70)=5.76$, $p<.001$). Participants also evaluated the face paired with the neutral biography more negatively than the face paired with the positive biography, though this difference was only marginally significant, $t(70)=1.83$, $p=.07$.

Learning generalization

As expected, there was a main effect of concurrently presented behavior on participants' evaluation of the new (morph) faces (Table 1), $F(3, 210)=80.90$, $p<.001$, $\eta_p^2=.54$. Participants evaluated faces presented with negative behaviors ($M=3.98$, $SD=1.21$) more negatively than those presented with neutral ($M=5.32$, $SD=1.21$), $t(70)=7.63$, $p<.001$) or positive behaviors ($M=6.20$, $SD=1.15$; $t(70)=11.67$, $p<.001$), and those with neutral behaviors more negatively than those with positive behaviors, $t(70)=9.26$, $p<.001$. Participants also evaluated faces presented with negative behaviors more negatively than faces presented without behaviors ($M=5.24$, $SD=1.08$; $t(70)=8.01$, $p<.001$) and those without behaviors more negatively than those with positive behaviors ($t(70)=8.12$, $p<.001$). They did not evaluate faces presented without behaviors differently from those presented with neutral behaviors ($t<1$).

Importantly, learning generalization also occurred, as evidenced by a significant effect of valence of learning (Table 1), $F(2, 140)=6.63$, $p=.002$, $\eta_p^2=.09$. Participants evaluated faces that were similar to the face paired with the negative biography ($M=4.98$, $SD=1.06$) more negatively than those that were similar to the face paired with the neutral ($M=5.26$, $SD=.94$; $t(70)=2.89$, $p=.005$) or positive biography ($M=5.31$, $SD=1.04$; $t(70)=3.06$, $p=.003$). Participants also evaluated faces that were similar to the neutral face more negatively than those that were similar to the positive face, though this difference was not significant, $t<1$. The interaction between concurrently presented behavior and valence of learning was not significant ($F<1$), indicating that learning generalization occurred regardless of concurrent behavioral information.

Discussion

As expected, there was a large effect of concurrently presented behavioral information on evaluation of the new (morph) faces. For example, new (morph) faces that were presented with negative behaviors were evaluated more negatively than those presented with no behaviors, neutral behaviors, or positive behaviors. Critically, there was also an independent effect of valence of learning. Those faces that were similar to the familiar negative face were evaluated more negatively than those that were similar to the familiar positive face. Thus, learning generalization persisted in the presence of relevant behavioral information, suggesting that people continue to rely on physical similarity even as they learn more information about a person.

Table 1
Mean (and standard deviation) trustworthiness ratings of new (morph) faces in Experiment 1 that were presented alone or presented with negative, neutral, or positive behaviors. The new (morph) faces contained 50% of the negative, neutral, or positive familiar faces. The main effect of concurrently presented behavior is represented in the rows of the table, and the main effect of the valence of the familiar face is represented in the columns.

Concurrent behavior	Similar to negative	Similar to neutral	Similar to positive
No behavior	4.98 (1.46)	5.35 (1.27)	5.40 (1.29)
Negative behavior	3.86 (1.57)	4.06 (1.49)	4.02 (1.47)
Neutral behavior	5.06 (1.50)	5.44 (1.37)	5.46 (1.32)
Positive behavior	6.04 (1.43)	6.21 (1.28)	6.36 (1.41)

While physical similarity continued to influence impressions, the effect of physical similarity on evaluation was smaller than the effect of concurrent behavioral information. Although it is tempting to draw conclusions about the relative importance of these different sources of information from this disparity, it is also worth keeping in mind that the relative size of the effects depends on the type of learning and the stimuli used. For instance, compared to the learning that occurs in the real world, exposure to the three familiar faces was quite brief. At the same time, the behavioral information was strongly valenced. Given the range of possible learning procedures and stimuli, obtaining an accurate estimate of the effect sizes for physical similarity and other types of information would prove difficult.

We found that physical similarity to familiar others influenced face evaluation in the presence of written behaviors, and we believe that this would hold true in the presence of actual behaviors. First, although we have examined the influence of physical similarity on face evaluation, it is worth keeping in mind that physical similarity has also been found to influence behavior: in particular, Lewicki (1985) found that the nature of participants' interaction with an initial experimenter influenced whether they chose to interact with a second experimenter who resembled the first. Second, Andersen and colleagues have found that similarity to significant others along non-physical dimensions influences a range of outcome measures. In one such study, Berk and Andersen (2000) even went so far as to show that participants' beliefs about similarity could influence the behavior of their naïve interaction partners: when participants believed the person they were talking to on the phone resembled a positive versus negative significant other, that person expressed more positive affect in their behavior. The findings of these studies suggest that physical similarity could continue to influence people's impressions of others during social interactions.

Experiment 2

Like all tasks, the evaluation of novel faces can be seen as relying on a combination of automatic and controlled processes: for instance, the evaluation of a novel face that resembles a familiar face may reflect both the automatic and intentional use of similarity information. While the first experiment examined the joint contributions of behavior and similarity to face evaluation, the second experiment moved on to investigate the extent to which participants can control their use of similarity in their evaluation of novel faces. Rather than relying on participants' self-reports of how similarity affected their judgments (e.g., Hill et al., 1990; Lewicki, 1985), we modified the learning generalization paradigm by introducing two manipulations during the evaluation of the new (morph) faces.

[1] In the post-rating phase of the first experiment, 11 participants rated the face paired with the positive biography more negatively than the face paired with the negative biography. Of these 11 participants, six failed to show evidence of learning for both the negative and positive faces when comparing pre- and post-ratings, and four failed to show learning for either the negative face or the positive face.

First, we introduced an instructional manipulation based on the logic of the process dissociation procedure (Jacoby, 1991). In this procedure, control over a behavior is defined as the difference in performance when automatic processes work in accord with controlled processes (facilitation) versus when they work against them (interference). This procedure was originally developed as a way to separate the contributions of recollection and familiarity to recognition memory. In particular, Jacoby (1991) compared performance on a task where familiarity facilitated correct responding (participants were asked to respond "old" to items seen at any point in the experiment) to performance on a task where familiarity interfered with correct responding (participants were asked to respond "old" only to items seen during a particular part of the experiment) and then estimated the relative contributions of recollection and familiarity to performance. While this procedure was developed in the context of recognition memory for words, more recently it has been used to address questions in social psychology (Payne, 2001; for a review, see Payne, 2008).

Bringing the logic of the process dissociation procedure to bear on the current study, we manipulated participants' intentions regarding use of similarity during the face evaluation task. Participants were either given no instructions regarding similarity, as in the previous experiment, or they were told that they would be seeing a series of faces that were similar to the familiar faces and shown an example morph. Participants who were told about similarity were either asked to take this information into account when making their judgments (inclusion condition) or they were asked to disregard this information (exclusion condition). If participants are able to intentionally increase their use of similarity information, we should see increased learning generalization in the inclusion condition relative to baseline. Meanwhile, if participants are able to control their use of similarity information, we should see decreased learning generalization in the exclusion condition relative to baseline.

Previous work has used the process dissociation procedure as a way of examining whether experimental manipulations influence the automatic or controlled contributions to a process. For instance, Ferreira, Garcia-Marques, Sherman, and Sherman (2006) found that manipulations including cognitive load and priming had predictable effects on the automatic and controlled contributions to judgment under uncertainty. More recently, McCarthy and Skowronski (2011) demonstrated that manipulating processing goals influenced the automatic but not controlled contribution to spontaneous trait inferences, while a temporal delay between learning and recognition had the opposite effect and influenced the controlled contribution.

In line with these studies, we introduced a cognitive load manipulation to our experiment. On half of the face evaluation trials, a 6-digit number appeared before the face. On these trials, participants were asked to indicate whether a second number that appeared after the face was the same or different from the first number. In the past, cognitive load manipulations have been used to demonstrate that processes are efficient, meaning that they can occur even when attention is directed elsewhere (Bargh, 1994). The cognitive load procedure used in the current study was adapted from Todorov and Uleman (2003), where it was used to show the efficiency of binding inferred traits from behavioral information to faces. Here, we introduced a load manipulation to explore the extent to which detection of similarity to familiar faces and use of this information in evaluation requires cognitive resources.

Examining the influence of load on learning generalization in the baseline condition allows us to test the efficiency of use of similarity information. To the extent that detecting similarity to familiar faces and using this information requires cognitive resources, load should diminish learning generalization. Further, examining the influence of load on learning generalization in the instruction conditions allows us to test the efficiency of using or disregarding similarity information. To the extent that intentionally using similarity information requires cognitive resources, load should diminish learning generalization in the inclusion condition. On the other hand, to the extent that disregarding similarity information requires cognitive resources, load should have the opposite effect on learning generalization in the exclusion condition and should increase the use of similarity information.

Method

Participants

Two hundred fifty-seven undergraduate students and members of the Princeton University community participated in this experiment for partial course credit or payment.

Stimuli

Stimuli were forty-five faces from the set of photographs used in Experiment 1 (Goldstone & Steyvers, 2001). Nine photographs of young to middle-age men were selected to be used as the familiar faces. These nine faces were divided into three groups and participants learned to associate each group with negative, neutral, or positive behaviors. The behaviors were from the database used in Experiment 1 (Fuhrman et al., 1989). The face and behavior groups were counterbalanced across participants.

Each familiar face was morphed with four novel faces, making a total of 36 new (morph) faces. Faces were morphed to a level of 50% using code from Steyvers (1999). For each set of four morphs, two of the morphs were presented in the no load condition and two in the load condition. The faces presented in each condition were counterbalanced across participants. For the cognitive load trials, we generated 18 random six-digit numbers using the website random.org. For nine of these trials, the second number matched the first. For the remaining nine trials, we changed one of the digits in the second number. To increase task difficulty, the digit we changed was never the first.

Procedures

Learning. Prior to seeing the nine faces paired with behaviors, participants were asked to evaluate the trustworthiness of the faces based on appearance alone. As in Experiment 1, participants were told that we were interested in first impressions and that there were no right or wrong answers. Trials began with a 500-millisecond fixation cross and faces remained on the screen until participants responded using the number keys 1 (not at all trustworthy) to 9 (extremely trustworthy). After evaluating the faces, participants followed the learning procedure described in Verosky and Todorov (2010). Participants saw each face paired with negative, neutral, or positive behaviors. Participants were asked to form impressions of the people by imagining them actually performing the behaviors. The faces were blocked so that participants saw each face paired with a first behavior before seeing any of them paired with a second behavior. Faces were always paired with the same type (negative, neutral, or positive) behavior. Each trial began with a 1000-millisecond fixation cross. The presentation of faces and behaviors was self-paced: face-behavior pairs remained on the screen until participants pressed the spacebar. During each block, the faces appeared in a different random order. After seeing each face paired with three different behaviors, participants saw each face alone and were asked to indicate whether it was previously shown with negative, neutral, or positive behaviors. The word "correct" or "incorrect" appeared for 1000 ms after each response. If participants gave an incorrect response, they saw each of the nine faces paired with additional behavior of the same valence before completing another test round. The task ended when participants reached 100% correct or when they completed eight test rounds, whichever came first. Each face was paired with five different behaviors before the face-behavior pairs repeated.

Face evaluation. Participants in the similarity inclusion and similarity exclusion conditions were told that they would be evaluating a series of morphed faces containing small percentages of the faces from the learning task. They were shown an example of a new (morph) face in the center of the screen, along with the familiar and novel faces used to create it. This new (morph) face served as an example of what morphed faces looked like, and it was not shown again.

Participants in all conditions were told that we were interested in their ability to process two different types of information at the same time. On no load trials, participants were asked to evaluate faces according to how trustworthy they thought they were. On the cognitive load trials, participants were asked to hold 6-digit numbers in memory while evaluating the faces. On these trials, a number appeared prior to the face. After evaluating the face, participants were asked to indicate whether a second number was the same as or different from the first number.

As in the previous experiments, participants in the baseline condition were told that we were interested in first impressions and that there were no right or wrong answers. Participants in the similarity *inclusion* condition were asked to take into account the similarity between the new (morph) face and the familiar faces from the previous phase of the task. They were given the example that if a face was similar to a face that was previously associated with positive behaviors, then they should rate the face more positively than they would otherwise. Finally, participants in the similarity *exclusion* condition were asked not to take into account the similarity between the new (morph) face and the familiar faces when making their ratings. They were given the example that if a face was similar to a face that was previously associated with positive behaviors, they should not rate the face more positively than they would otherwise.

Participants completed ten practice trials to familiarize them with the procedure. Five of the practice trials were load trials; all of the faces used in the practice were novel. Each no load trial began with a 500-millisecond fixation cross. Then, a face was presented for 500 ms. After the face, a rating scale appeared and remained on the screen until participants responded using the number keys 1 (not at all trustworthy) to 9 (extremely trustworthy). Each load trial began with a number, which remained in the center of the screen for 2 s. After the presentation of the number, load trials followed the same procedure as no load trials. After the face and rating, a second number appeared, along with the prompt “Is this number the same as or different from the original number?” The number remained on the screen until participants responded using the letter keys g (same) and h (different). The inter-trial interval was 2s.

After completing the practice trials, participants evaluated the 36 new (morph) faces. The order of load and no load trials was randomized for each participant. Finally, participants were told that they would see faces from the learning task. Participants evaluated these nine faces following the procedure from the beginning of the learning task[2].

The overall design was a 3 (instructions: none, include similarity, exclude similarity) × 3 (valence of learning: negative, neutral, or positive) × cognitive load (no load, load) mixed ANOVA, with the first factor between subjects and the last two within-subject.

Results

Manipulation checks

Half of participants were able to identify the valence of behavior (negative, neutral, or positive) paired with each face after two rest rounds ($M = 50\%$), 75% were able to do so after four test rounds ($M = 75\%$), and nearly 95% were able to do so after eight test rounds ($M = 93\%$). Seventeen participants (6.6% of the total number) failed to learn the face-valence associations after eight test rounds and their data were excluded from further analysis.

There was a significant effect of the valence of the learned behavioral information on participants' evaluation of the familiar faces, $F(2, 478) = 442.02$, $p < .001$, $\eta_p^2 = .65$. Participants evaluated the faces previously paired with negative behaviors ($M = 2.92$, $SD = 1.39$) significantly more negatively than those previously paired with neutral ($M = 4.99$, $SD = 1.00$; $t(239) = 20.32$, $p < .001$) or positive behaviors ($M = 6.69$, $SD = 1.59$; $t(239) = 23.27$, $p < .001$). Participants also rated faces previously paired with neutral behaviors significantly more negatively than those paired with positive behaviors, $t(239) = 15.70$, $p < .001$.

Participants were able to correctly indicate whether a number was the same as or different from the original number 91% of the time ($SD = 10\%$). Accuracy was significantly higher than chance (chance = 50%; $t(239) = 64.56$, $p < .001$), but significantly lower than perfect performance ($t(239) = 13.83$, $p < .001$).

Learning generalization

There was a significant main effect of valence of learning (Table 2; $F(2, 474) = 111.95$, $p < .001$, $\eta_p^2 = .32$), with participants evaluating faces that were similar to familiar negative faces ($M = 4.69$, $SD = .88$) more negatively than those that were similar to familiar neutral ($M = 5.08$, $SD = .80$; $t(239) = 7.39$, $p < .001$) or familiar positive faces ($M = 5.43$, $SD = .89$; $t(239) = 10.39$, $p < .001$). In turn, participants evaluated faces that were similar to familiar neutral learned faces more negatively than those that were similar to familiar positive faces ($t(239) = 6.67$, $p < .001$).

There was no main effect of instruction condition on face evaluation ($F(2, 237) = 1.71$, $p = .18$, $\eta_p^2 = .01$), but there was a significant interaction between instructions and valence of learning ($F(4, 474) = 43.85$, $p < .001$, $\eta_p^2 = .27$). In particular, there was a stronger effect of valence of learning in the inclusion condition as compared to the baseline ($F(2, 237) = 57.42$, $p < .001$, $\eta_p^2 = .27$) and exclusion conditions ($F(2, 308) = 60.99$, $p < .001$, $\eta_p^2 = .28$), indicating that participants were able to increase their use of similarity information when asked to do so. On the other hand, the effect of valence of learning did not significantly differ between the baseline and exclusion conditions ($F < 1$), indicating that participants were not able to completely exclude similarity from their judgments.

There was also a main effect of load ($F(1, 237) = 8.94$, $p = .003$, $\eta_p^2 = .04$), with participants evaluating faces more positively when they were under load ($M = 5.11$, $SD = .72$) versus not under load ($M = 5.02$, $SD = .71$). This main effect was qualified by a significant interaction with

Table 2
Mean (and standard deviation) trustworthiness ratings of new (morph) faces in Experiment 2 as a function of instructions and cognitive load. Participants were not told about the similarity between the morphed faces and the learned faces (baseline), or they were asked to either take similarity into account (inclusion) or to avoid taking it into account (exclusion). The new (morph) faces contained 50% of the negative, neutral, or positive familiar faces. The main effects of instructions, load, and valence of the familiar face are represented in the rows of the table containing the instruction condition, the rows containing the load condition, and the columns, respectively. The means for the main effects reported in the text are weighted by the number of participants per instruction condition.

		Similar to negative	Similar to neutral	Similar to positive
Baseline	No load	5.00 (.90)	5.11 (.95)	5.35 (.97)
	Load	5.05 (.91)	5.25 (.89)	5.29 (.91)
Inclusion	No load	4.11 (.93)	4.91 (.84)	5.96 (.97)
	Load	4.18 (.96)	5.11 (.95)	5.78 (1.10)
Exclusion	No load	4.76 (.81)	4.90 (.83)	5.07 (.91)
	Load	4.95 (.82)	5.18 (.84)	5.17 (.83)

[2] In the post-ratings phase of the experiment, 15 participants failed to rate the negative learned faces more negatively than the positive learned faces. Of these 15 participants, one failed to show evidence of learning for both the negative and positive faces when comparing pre- and post-ratings, and ten failed to show evidence of learning for either the negative or the positive faces. Thus as in Experiment 1, using the post-ratings alone provided a means of collapsing across learning for negative and positive faces.

valence of learning ($F(2, 474) = 5.72$, $p = .003$, $\eta_p^2 = .02$), such that the main effect of valence of learning was slightly diminished when participant were under load ($F(2, 474) = 59.77$, $p < .001$, $\eta_p^2 = .20$) versus not under load ($F(2, 474) = 89.39$, $p < .001$, $\eta_p^2 = .27$).

Finally, the interaction between load and instruction condition was not significant, $F(2, 237) = 2.98$, $p = .053$, $\eta_p^2 = .02$, nor was the three-way interaction between load, valence of learning, and instruction condition, $F < 1$. However, planned comparisons revealed a more pronounced effect of load on valence of learning in the inclusion condition ($F(2, 150) = 3.23$, $p = .04$, $\eta_p^2 = .04$) versus the baseline ($F(2, 166) = 1.25$, $p = .29$, $\eta_p^2 = .01$) and exclusion conditions ($F(2, 158) = 1.29$, $p = .28$, $\eta_p^2 = .02$)[3].

Discussion

As in the previous experiment, participants evaluated faces that were similar to familiar negative faces more negatively than faces that were similar to familiar positive faces. Importantly, the instructions that participants were given regarding use of similarity information had a large effect on the magnitude of learning generalization. Specifically, instructions to take similarity into account led to a stronger learning generalization in the inclusion condition than in the baseline and exclusion conditions, indicating that participants were able to intentionally use similarity information when asked to do so. However, at the same time, instructions to avoid taking similarity into account did not eliminate learning generalization entirely, indicating that participants were not able to completely control their use of similarity information. Learning generalization in the absence of instructions looked very similar to learning generalization when participants were asked to disregard similarity information, suggesting that participants' default state may be more like the exclusion than inclusion condition.

In contrast to the manipulation of instructions, the manipulation of load only had a slight influence on learning generalization, suggesting that learning generalization is a relatively efficient process. In particular, although cognitive load decreased use of similarity information across all conditions, this effect was very small and did not eliminate the use of similarity. In other words, even when participants were busy rehearsing a six-digit number, they still evaluated faces that were similar to familiar negative faces more negatively than those that were similar to familiar positive faces.

Finally, an inspection of the effect of load as a function of instructions revealed that load affected learning generalization in the inclusion condition, but not in the baseline or exclusion conditions. The finding that load interfered with use of similarity information in the inclusion condition provides further evidence that what participants were able to control was *increased* use of similarity information. At the same time, the fact that load did not interfere with learning generalization in the baseline and exclusion conditions suggests that participants in these conditions were simply unable to remove similarity information from their judgments, even when they tried to do so.

[3] As an alternative way of examining the face evaluation data, we estimated the automatic and controlled contributions to learning generalization using the process dissociation procedure (Jacoby, 1991). Unfortunately, because the instruction conditions were between subjects in order to avoid any effect of order on ability to control similarity, we could not conduct statistical tests on the automatic and controlled estimates. We could only describe them at the condition level. For the purposes of calculating the automatic and controlled estimates, we defined the effect of valence of learning as the difference in evaluation between the new (morph) faces that were similar to the familiar positive and familiar negative faces. After transforming the continuous ratings into proportions by dividing by the maximum value on the rating scale, we found a higher controlled ($C = .17$) than automatic contribution ($A = .04$) when participants were not under load, as would be expected given the stronger learning generalization in the inclusion versus exclusion condition. We found slightly lower controlled ($C = .15$) and automatic ($A = .03$) contributions when participants were under load, indicating that load decreased both controlled and automatic use of similarity information.

General discussion

Physical similarity to familiar others has been found to influence impressions of others across a range of settings. For example, physical similarity to the self can influence trust behavior in economic games (DeBruine, 2002), physical similarity to significant others can lead participants to infer that future interaction partners possess other characteristics of their significant other (Kraus & Chen, 2010), and physical similarity to famous people can influence judgments of trustworthiness (Tanner & Maeng, 2012). Meanwhile, at the level of the group, greater familiarity of own-race than other-race faces, which likely reflects similarity to a shared facial prototype, can partially explain ingroup favoritism (Zebrowitz et al., 2007, 2008). While these existing studies can be seen as examining the different types of learning that generalize (e.g., learning about the self, about significant others, etc.), the current experiments contribute to the literature by examining the mechanisms underlying such effects.

In two experiments, participants evaluated faces that were similar to familiar negative faces more negatively than those that were similar to familiar positive faces. In Experiment 1, learning generalization persisted even in the presence of relevant behavioral information. In Experiment 2, learning generalization occurred regardless of the instructions given to participants about use of similarity: while participants were able to increase their use of similarity information, they were not able to exclude similarity from their judgments. In addition, although learning generalization was diminished during a demanding secondary task, it was not eliminated altogether.

Our findings speak to the broad scope of learning generalization. In particular, the persistence of learning generalization in Experiment 1 shows that physical similarity is not simply something people only rely on when they don't have other information available, but rather is part of a more general impression formation process. Meanwhile, participants' inability to completely exclude similarity information from their judgments in Experiment 2 suggests that physical similarity is difficult to ignore. Thus, learning generalization is likely to occur even a) under circumstances when you know other information about a person and b) despite intentions to ignore it.

Although we did not directly assess whether participants intended to use similarity information, both experiments address this issue indirectly. In Experiment 1, insofar as similarity information is less diagnostic of character than information about behavior, one might imagine that participants may change their intentions regarding use of similarity as other, more diagnostic information becomes available; if this is the case, the persistence of learning generalization in the presence of relevant behavioral information suggests that use of similarity information may not be entirely intentional. Meanwhile, in Experiment 2, the fact that learning generalization was very similar in the exclusion and baseline conditions suggests that participants were not intentionally including similarity information in judgments.

A related question is whether people need to consciously recognize physical similarity in order for it to influence evaluation. While previous studies have found effects of physical similarity without participants detecting the manipulation of interest (e.g., Günaydin et al., 2012; Hill et al., 1990; Lewicki, 1985), a recent study by Tanner and Maeng (2012) suggests a more nuanced answer to this question. While learning in our study took place in the lab, these authors capitalized on learning that different people in a culture have in common by using famous faces. In one experiment, they found that participants preferred a morphed face containing 35% of George Bush's face to a matched control, irrespective of self-reported liking for Bush. However, in another experiment, they found that a preference for a morph containing 35% of Tiger Woods' face reversed itself after the Tiger Woods scandal. The authors reconcile these findings by suggesting that participants implicitly recognized the familiar faces at an intermediate level of specificity: although participants did not report recognizing the celebrity faces used in the morphs, they were able to access an overall valence evaluation of the person in question. In the case of Bush, this overall evaluation was positive, possibly because of his association with the role of president of the United States; meanwhile, in the case of Tiger Woods, the evaluation reversed itself. Although the idea of an intermediate level of implicit recognition is intriguing, many questions, including whether recognition occurs along dimensions other than valence, remain.

Taking a step back, it is possible that the mechanisms underlying use of physical similarity in face evaluation may differ depending on the type of learning, for instance whether the learning takes place in regard to a significant other or social group. However, even if this is true, it seems likely that the current findings will be relevant for these other cases as well: the learning in this study is minimal and the similarities subtle compared to the types of learning we engage in on a day-to-day basis.

In conclusion, while Verosky and Todorov (2010) demonstrated that affective learning can generalize to perceptually similar novel faces, the current studies investigate the mechanism underlying such learning generalization. Experiment 1 demonstrated that people continue to use physical similarity in their judgments of faces, even when there is behavioral information available. Meanwhile, Experiment 2 showed that while participants are able to increase their use of similarity information, they are not able to disregard it. Thus, in line with the idea that similarity is fundamental to face processing, we find evidence that learning generalization is a powerful and relatively automatic process, which influences the evaluation of faces under a variety of circumstances.

Appendix A. Biographical vignettes

Negative: He recently finished a prison sentence for attempted manslaughter. Since he got out, he has been working at his uncle's deli, but he has a bad temper and he is very unpopular with his co-workers, so his uncle has asked him to find another job. He spends his spare time in the local bars, getting drunk and picking fights with strangers.

Neutral: He has a steady job. He doesn't find it particularly interesting, but it does provide him with a steady source of income. He recently joined a bowling league with some colleagues from work and they meet up every Tuesday night. In his spare time, he practices in hopes of improving his game.

Positive: He runs a very successful nonprofit business that helps fund afterschool programs for disadvantaged children. He is also a volunteer firefighter and he has risked his own life a number of times in order to save others. In his spare time, he enjoys spending time with his children, often taking them to the zoo or to ballgames.

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.

Bargh, J. A. (1994). The four horsemen of automaticity: Intention, awareness, efficiency, and control as separate issues. In R. S. Wyer, & T. K. Srull (Eds.), *Handbook of social cognition: Vol. 1, Basic processes* (pp. 1–40) (2nd ed.) . Hillsdale, NJ: Erlbaum.

Baron, S. G., Gobbini, M. I., Engell, A. D., & Todorov, A. (2011). Amygdala and dorsomedial prefrontal cortex responses to appearance-based and behavior-based person impressions. *Social Cognitive and Affective Neuroscience*, *6*, 572–581.

Berk, M. S., & Andersen, S. M. (2000). The impact of past relationships on interpersonal behavior: Behavioral confirmation in the social–cognitive process of transference. *Journal of Personality and Social Psychology*, *79*, 546–562.

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.

Carlston, D. E., & Skowronski, J. J. (1994). Saving in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, *66*, 840–856.

Carlston, D. E., Skowronski, J. J., & Sparks, C. (1995). Savings in relearning II: On the formation of behavior-based trait associations and inferences. *Journal of Personality and Social Psychology*, *69*, 420–436.

Crawford, M. T., Skowronski, J. J., Stiff, C., & Scherer, C. R. (2007). Interfering with inferential, but not associative processes underlying spontaneous trait inference. *Personality and Social Psychology Bulletin*, *33*, 677–690.

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

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

Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Sherman, J. W. (2006). Automatic and controlled components of judgment and decision making. *Journal of Personality and Social Psychology*, *91*, 797–813.

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.

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology. General*, *130*, 116–139.

Günaydin, G., Zayas, V., Selcuk, E., & Hazan, C. (2012). I like you but I don't know why: Objective facial resemblance to significant others influences snap judgments. *Journal of Experimental Social Psychology*, *48*, 350–353.

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.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.

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.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126*, 390–423.

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

McCarthy, R. J., & Skowronski, J. J. (2011). The interplay of controlled and automatic processing in the expression of spontaneously inferred traits: A PDP analysis. *Journal of Personality and Social Psychology*, *100*, 229–240.

Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior*, *27*, 237–254.

O'Toole, A. J. (2011). Cognitive and computational approaches to face recognition. In A. Calder, J. V. Haxby, M. Johnson, & G. Rhodes (Eds.), *Handbook of face perception* (pp. 15–30). Oxford: Oxford University Press.

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

Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion*, *9*, 128–133.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*, 181–192.

Payne, B. K. (2008). What mistakes disclose: A process dissociation approach to automatic and controlled processes in social psychology. *Social and Personality Psychology Compass*, *2*, 1073–1092.

Quine, W. V. (1969). Natural kinds. In W. V. Quine (Ed.), *Ontological relativity and other essays* (pp. 114–138). New York: Columbia University Press.

Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research*, *46*, 2977–2987.

Rhodes, G., & Leopold, D. A. (2011). Adaptive norm-based coding of face identity. In A. Calder, J. V. Haxby, M. Johnson, & G. Rhodes (Eds.), *Handbook of face perception* (pp. 263–286). Oxford: Oxford University Press.

Rudoy, J. D., & Paller, K. A. (2009). Who can you trust? Behavioral and neural difference between perceptual and memory-based influences. *Frontiers in Human Neuroscience*, *3*, 16.

Said, C., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, *9*, 260–264.

Skowronski, J. J., Carlston, D. E., Mae, L., & Crawford, M. T. (1998). Spontaneous trait transference: Communicators take on the qualities they describe in others. *Journal of Personality and Social Psychology*, *74*, 837–848.

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

Tanner, R. J., & Maeng, A. (2012). A tiger and a president: Imperceptible celebrity facial cues influence trust and preference. *Journal of Consumer Research*, *39*, 769–783.

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. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*, 1051–1065.

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

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*, 482–493.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *43*, 161–204.

Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science*, *21*, 779–785.

Zebrowitz, L. A. (1996). Physical appearance as a basis of stereotyping. In N. MacRate, 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., Fellous, J., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. *Personality and Social Psychology Review*, *7*, 194–215.

Zebrowitz, L. A., Kikuchi, M., & Fellous, J. M. (2010). Facial resemblance to emotions: Group differences, impression effects, and race stereotypes. *Journal of Personality and Social Psychology*, *98*, 175–189.

Zebrowitz, L. A., & McDonald, S. M. (1991). The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior*, *15*, 603–623.

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.