

**From the holdings of the Department of Special Collections , Princeton University Library,
Princeton, NJ**

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyright law. We do not charge any permission or usage fees for the publication of images of material in our collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff. For more information regarding copyright, credit, and citation please review the information on the Copyright, Credit and Citation Guidelines page of our website.

<https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines>. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. <https://library.princeton.edu/special-collections/ask-us>

Doctoral Dissertations, 2011-2026

Psychology

2012

More than a face: Interactions between visual and non-visual social knowledge

Verosky, Sara

Princeton, NJ : Princeton University

<http://arks.princeton.edu/ark:/88435/dsp01r781wg05s>

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. <https://www.copyright.gov/title17/title17.pdf>

**MORE THAN A FACE: INTERACTIONS BETWEEN VISUAL AND
NON-VISUAL SOCIAL KNOWLEDGE**

Sara Charlotte Verosky

**A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY**

**RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF PSYCHOLOGY**

Advisor: Alexander Todorov

June 2012

© Copyright by Sara Charlotte Verosky, 2012. All rights reserved.

Abstract

Although we represent other people in terms of both what they look like and how they behave, these two aspects of person knowledge have rarely been studied together. While vision scientists have focused on the processing of faces and bodies, social psychologists have focused on the cognitive construal of others. In this dissertation, I present a series of experiments that aim to better understand how we think about other people by considering how we combine visual and non-visual knowledge in our person representations. In particular, these experiments focus on one type of interaction between visual and non-visual social knowledge: the influence of social knowledge on face perception. In the first part of the dissertation, I demonstrate that rapidly learned affective associations generalize to perceptually similar but novel faces. In the second part, I examine the mechanisms underlying this type of learning generalization. In the last part of the dissertation, I present evidence from fMRI that different facial identities are associated with unique patterns of neural activity in ventral visual cortex and I examine the influence of social knowledge on these representations. Together, these data suggest that what we know about others can change the way we see their faces.

Table of Contents

Abstract.....	iii
Table of Contents.....	iv
Acknowledgements.....	v
Introduction.....	1
Part 1: Generalization of Affective Learning About Faces to Perceptually Similar Faces.....	7
Part 2: Mechanisms Underlying Affective Learning Generalization.....	20
Part 3: Representations of Individuals in Ventral Temporal Cortex Defined by Faces and Biographies.....	58
Conclusions and Future Directions.....	83
References.....	86
Footnotes.....	98
Supplementary Material for Part 1.....	99
Supplementary Material for Part 3.....	103
Appendix A.....	111
Appendix B.....	112

Acknowledgements

First, thank you to my advisors Alex Todorov and Nick Turk-Browne for their guidance and support. Thank you Alex for your compassion, optimism, and insight, and for always encouraging me to think for myself; thank you Nick for welcoming me into your lab, your openness to ideas, and for pushing me to do more than I thought was possible. I hope you both know how incredibly grateful I am to have you as mentors. Also, thank you to the remaining members of my committee – Susan Fiske, Asif Ghazanfar, and Dan Osherson – for their thoughtful comments through the years.

Thank you to my family and friends for cheering me on through the ups and downs of graduate school. Thank you to my mom Cynthia for always being happy to discuss whatever quandary I found myself in, to my dad John for helping me to keep things in perspective, to my brother Niels for visiting, and to my grandmother Charlotte for our weekly conversations. Finally, thank you to all my Princeton friends, with special thanks to my cohort, my officemates, and Meg, Katie, and Abby, for being there in a hundred little (and big) ways.

Introduction

Although we represent other people in terms of both what they look like and how they behave, these two aspects of person knowledge have rarely been studied together. While vision scientists have focused on the processing of faces and bodies, social psychologists have focused on the cognitive construal of others. However, even though these two types of information have been studied separately, they are in fact closely related: From the perspective of vision science, it is important to keep in mind that we are social creatures, and the visual system itself was likely shaped by social constraints. Meanwhile, from the perspective of social psychology, it is important to note that we are visual creatures, and that much of our information about the world comes to us through vision. In this dissertation, I present a series of experiments that aim to better understand how we think about other people by considering how we combine these two types of knowledge in our person representations.

There are different ways to go about studying the interaction between visual perception and social knowledge. On the one hand, it is possible to ask how vision influences social knowledge: how do visual cues influence the types of social inferences people draw? On the other hand, it is also possible to ask how social knowledge influences visual perception: how does what people know influence what they perceive? In the experiments presented here, I take the latter approach and I examine how social knowledge influences the perception of faces. I focus on faces because they are among the most important social stimuli we encounter: they not only provide information about identity, but they convey a wealth of other information as well. For instance, a simple

glance at someone's face is enough to garner information about their gender, race, age, emotional state, and direction of attention.

People are very good at learning to associate faces with behavioral information (Carlston & Skowronski, 1994; Todorov & Uleman, 2002, 2003, 2004). Learning can be triggered by minimal information: simply seeing a face paired with a trait-implying behavior for 2 s is enough for participants to form corresponding trait associations (Todorov & Uleman, 2003). Moreover, these affective associations can persist independently of memory for the behavior that originally triggered them, as seen in work with healthy participants (Carlston & Skowronski, 1994; Todorov & Uleman, 2002) and patients with amnesia (Johnson, Kim, & Risse, 1985). Learning behavioral information also changes judgments based on facial appearance: reading behavioral descriptions influences ratings of physical attractiveness (Gross & Croften, 1977), face trustworthiness (Todorov & Olson, 2008), and of physical similarity between pairs of faces (Hassin & Trope, 2000). However, even though these findings demonstrate a powerful link between faces and behaviors, a question remains: are changes in evaluation of faces due to changes in the influence of conceptual knowledge or are they instead due to change in the perception of the faces themselves?

In the following experiments, I approach this question from two different angles. First, I examine how learning influences the evaluation of novel faces that resemble the learned faces: to the extent that learning changes perception of learned faces, we might expect it to influence the evaluation of perceptually similar novel faces as well. Second I use fMRI to examine how learning influences identity representations in ventral visual

cortex: if learning changes the perception of learned faces, we might expect to see changes in the representations in this part of the brain.

Overview of Experiments

Different people have different (looking) significant others, friends, and foes. The objective of the first part of this dissertation was to demonstrate that these different social face environments could shape individual face preferences. To test this idea, I had participants learn to associate faces with negative, neutral, or positive behaviors. Participants then evaluated new morphed faces that combined the novel faces with the learned faces. The morphs contain only small percentages of the learned faces and the categorization of the morphs as similar to the learned faces was not different from the categorization of actual novel faces. Nonetheless, the evaluation of the morphed faces was affected by their perceptual similarity to the familiar faces, such that participants evaluated morphs that were similar to negative learned faces more negatively than morphs that were similar to positive learned faces. This learning generalization effect suggests that general learning mechanisms based on similarity can account for idiosyncratic face preferences.

While the experiment in the first part of the dissertation serves as a laboratory demonstration of learning generalization, the three experiments in the second part adapt the paradigm developed above in order to investigate the processes by which physical similarity influences the evaluation of novel faces. Across these experiments, participants once again evaluated morphs that were similar to negative learned faces more negatively than morphs that were similar to positive learned faces. This learning generalization

effect was present when participants' judgments of the morphs were a) based not only on facial appearance but also on relevant behavioral information (Experiment 2.1); b) made under cognitive load (Experiment 2.3); and c) made under instructions not to use similarity information (Experiment 2.3). However, learning generalization was contingent on the ability to recognize the similarity of the morphs to the learned faces, although it was also evident for truly novel faces that participants perceived as similar to the learned faces (Experiment 2.2). 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.

In the last part of the dissertation, I use fMRI to examine how identity is represented in the brain and how learning changes these representations. Multiple regions in the fusiform gyrus respond more strongly to faces than to other categories of objects, but whether this response reflects the categorical detection of faces or the recognition of particular identities remains open to debate. Although univariate analyses in this region have revealed adaptation for repeated versus novel identities, the results of multivariate analyses have been less conclusive. In this experiment, I tested whether the creation of richer identity representations via training on visual and social information, and the use of an adaptation design, would reveal more robust representations of these identities in ventral temporal cortex. Examining the patterns of activation across voxels in bilateral fusiform gyri and right anterior temporal lobe, I identified unique patterns for particular identities. Attaching distinctive biographical information to identities did not increase the strength of these representations, but did produce a grouping effect whereby faces were represented more similarly to other faces associated with the same amount of such

information. Thus, identity exemplars are represented not only in anterior temporal lobe, but also in posterior visual areas thought to represent information at a categorical level. The suggestive influence of biographical information in shaping these representations provides additional evidence that 'visual' areas can be sensitive to other kinds of information, including from the social domain.

In sum, the experiments presented in this dissertation examine how social knowledge influences the visual perception of faces. While the first two parts of the dissertation show how social knowledge can shape our evaluation of novel faces, the last part moves on to examine how identities are represented in the brain and how social knowledge influences these representations.

Part 1: Generalization of Affective Learning About Faces to Perceptually Similar Faces

(Adapted from Verosky & Todorov, 2010)

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 these idiosyncratic “judge” contributions to judgments simply reflect random noise across judges. For example, people have more positive evaluations of faces that resemble themselves (Bailenson, Iyengar, Yee, & Collins, 2009; DeBruine, 2002, 2005). Thus, it seems that part of beauty really is in the “eye of the beholder.”

Consistent with the familiar face overgeneralization hypothesis (Zebrowitz, 1996; Zebrowitz & Collins, 1997), we argue that evaluation of novel faces is partially based on similarity to familiar faces. 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 eye glasses) when the interaction was pleasant, but chose the other person (who had long hair and no glasses) when the interaction was unpleasant.

The goal of the current study was to test whether rapidly learned affective associations with faces are generalized to novel faces that resemble the familiar faces. To the extent that evaluation of novel faces is influenced by their similarity to familiar faces, this would suggest that “unique” individual face preferences might result from a common underlying process of learning generalization.

To design the experiment, we capitalize on two phenomena: categorical perception of faces (Beale & Keil, 1995; Levin & Beale, 2000) and findings that social judgments of faces are easily changed when people are provided with information about the person (Bliss-Moreau, Barrett, & Wright, 2008; Carlston & Skowronki, 1994; 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 the face are attenuated, whereas differences between morphs on the category boundary are accentuated.

In the experiment, participants learned to associate faces with positive, neutral, or negative behavioral information. Then, they evaluated morphs of these faces and novel faces that were within the categorical boundary 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 one of the faces were identified as that face. Correspondingly, in the main study, participants evaluated novel faces that were morphed with 20% or 35% of the learned faces. In an additional study, we further showed that participants treated the morphs like completely novel faces. Specifically, after learning the associations of faces and behaviors, participants’ categorization of the morphs as

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

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

Methods

Participants

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

Stimuli

We used 54 photographs of men with neutral facial expressions from a set of black-and-white photographs of bald males (Kayser, 1997). We selected nine photographs of younger to middle-aged men to be used as the learned 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 chosen from a database of four hundred behaviors based on goodness ratings (Fuhrman, Bodenhausen, & Lichtenstein, 1989). The face and behavior groups were counterbalanced across participants.

Procedures

Preliminary study 1. To determine the location of the identity boundary between pairs of faces, we created a series of morphs. Participants ($N=16$) were asked to categorize each morph as looking like one or the other endpoint face (see Supplementary Material at the end of this dissertation). Based on the results (Figure 1.1B), we decided to use 20% and 35% morphs of the learned faces.

Learning. In the learning phase, participants were told that their task was to form person impressions. They were presented with 9 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 Bliss-Moreau et al. (2008) and Todorov and Uleman (2003). Each face was paired with three different behaviors of the same valence: negative, neutral, or positive. The faces were blocked so that participants saw all nine faces paired with a first behavior, then with a second, and then with a third. Each time, the faces were shown in a different random order. The presentation of the face-behavior pairs was self-paced and the inter-trial interval was 1000 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 negative, neutral, or positive behaviors. Participants received feedback about their accuracy. If a participant 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. This continued until the participant reached 100% correct or completed eight test rounds. Each face was paired with at most five different behaviors before the face-behavior pairs were repeated.

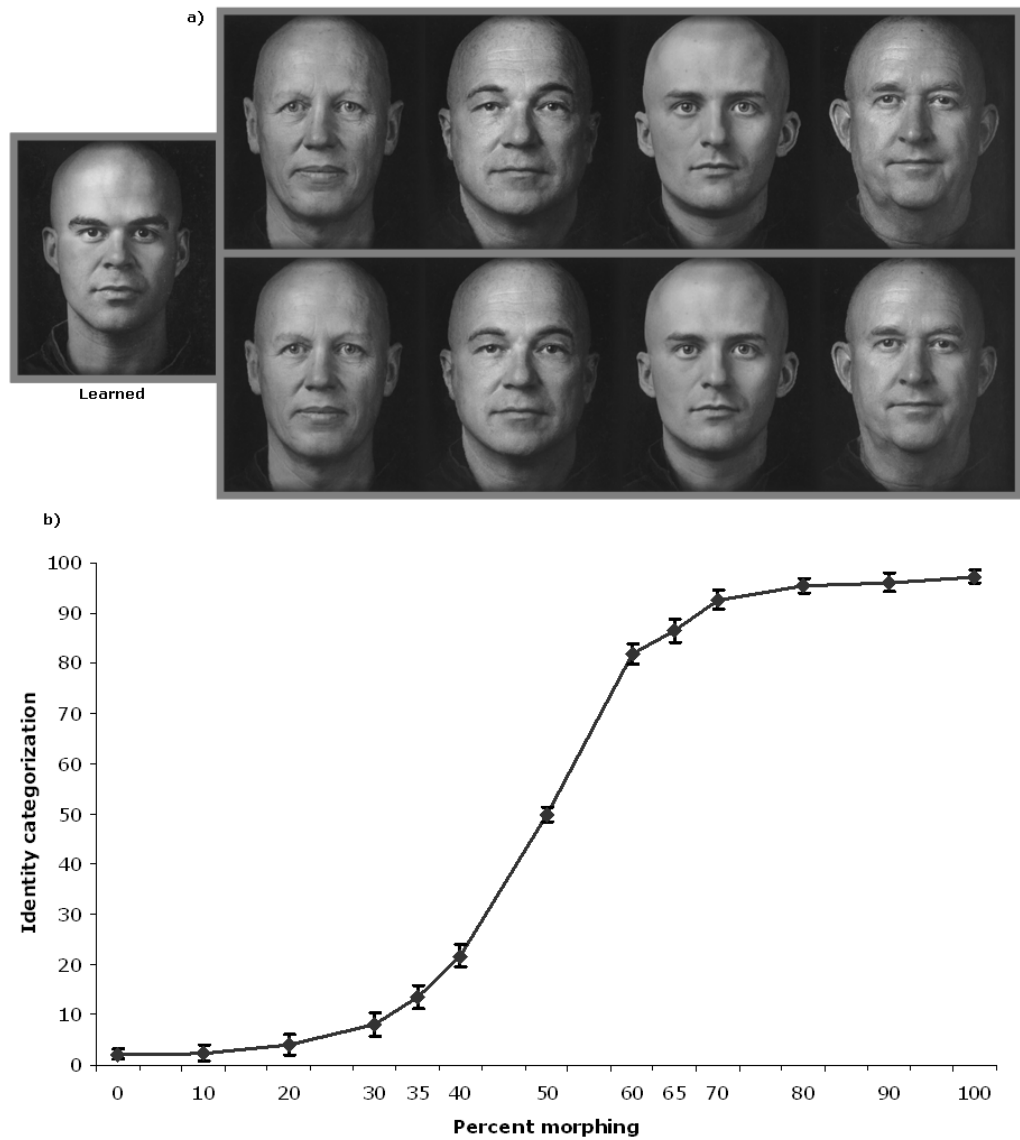


Figure 1.1. The category boundary between faces. (A) Example of stimuli used in the experiment. The top row shows morphs containing 20% of the learned face shown in the box on the left. The bottom row shows morphs containing 35% of the learned face. (B) Percent of the time participants ($N=16$) identified morphs as looking like the original face as a function of the percent of the original face present. Error bars represent standard error of the mean.

Preliminary study 2. After the learning phase, participants ($N=15$) 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. The morphed faces were the same as those used in the evaluation phase of the main experiment (see Supplementary Material).

Face evaluation. In the evaluation phase of the main experiment, participants were told that we were interested in first impressions and that they would evaluate both learned and novel faces. Prior work has demonstrated that trustworthiness judgments provide a good approximation of valence evaluation of faces (Oosterhof & Todorov, 2008) and, therefore, participants were asked to rate the faces on trustworthiness. Each of the 9 learned faces was morphed with 4 of 36 novel faces at two different levels of morphing (20% and 35%), using code from Steyvers (1999). Half of the novel faces were shown at one morphing level and half at the other, and these were counterbalanced across participants.

Each trial started with a 500-millisecond fixation cross. The face remained on the screen until the participant responded using the number keys from 1 (not at all trustworthy) to 9 (extremely trustworthy). Each participant rated all faces twice, and each time, the faces were shown in a different random order.

The design was a 3 (Valence of learning: positive vs. neutral vs. negative) X 2 (Morphing: 35% vs. 20% learned faces) repeated measures analysis of variance.

Results

Preliminary studies

At the 20% morphing level, the similarity between the morphs and the original face was difficult to detect (Figure 1.1A), and the morphs were almost always categorized as looking like the other endpoint face (Figure 1.1B). At the 35% morphing level, the morphs were categorized as looking like the latter face more than eighty-five percent of the time.

The additional preliminary study further showed that the morphs were treated as novel faces (Figure 1.2). On the majority of trials, participants categorized the morphs as novel faces ($M=.74$, $SD=.16$) and almost never categorized the morphs as learned faces ($M=.02$, $SD=.03$). These proportions were not significantly different from the respective proportions for actual novel faces, $t_s < 1$. Although participants sometimes categorized the morphs as similar to the learned faces ($M=.24$, $SD=.16$), this was not different from the categorization of novel faces as similar, $t < 1$ (see Supplementary Material).

Learning

In the learning phase of the main experiment, 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 test rounds. Comparing the judgments of the learned faces, there was 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 out of 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 54 participants who showed evidence for learning¹. 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: positive>neutral>negative. This was the case for only 35 participants.

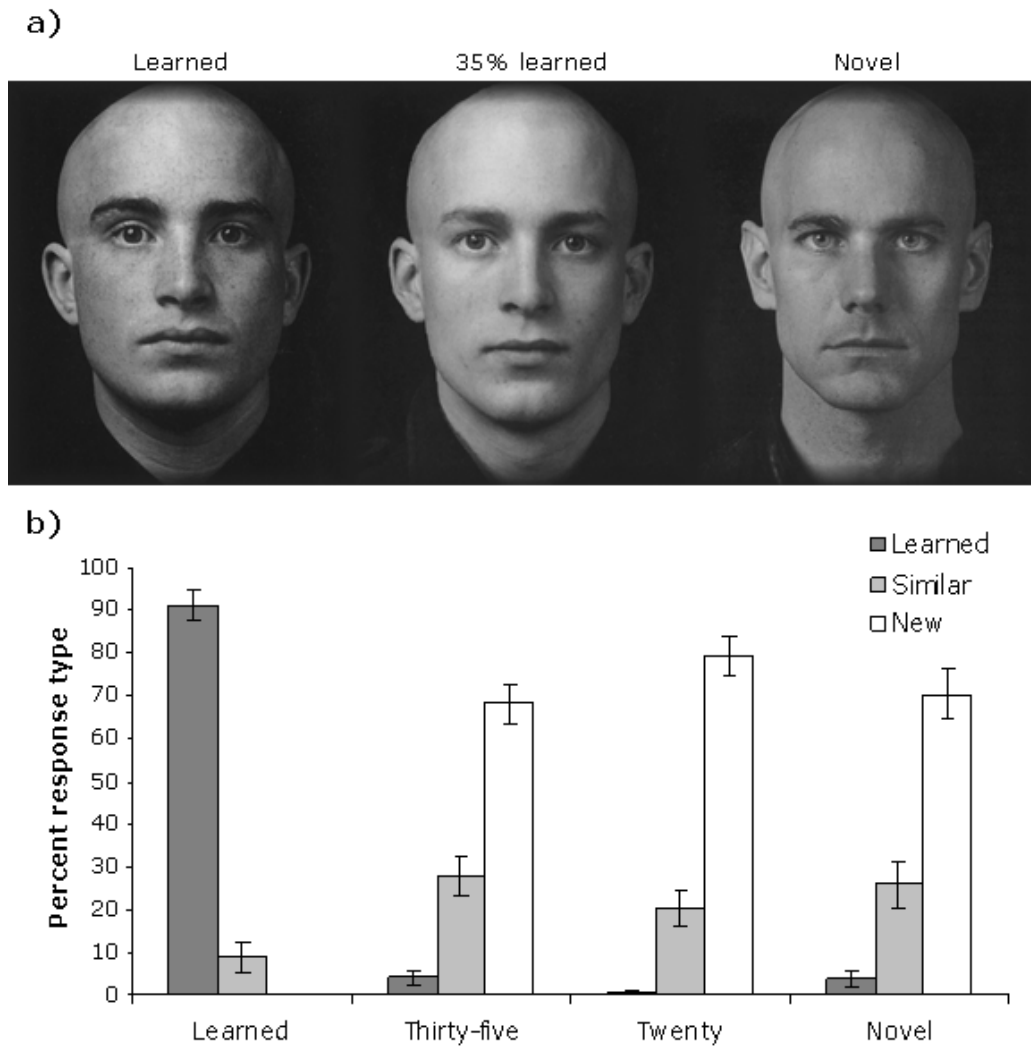


Figure 1.2. Similarity detection. (A) Example of a learned face, a 35% morph of this face, and a completely novel face. (B) Percent of the time participants categorized faces as learned, similar to learned, or new by type of face ($N=15$). Error bars represent standard error of the mean.

Learning generalization

As expected, participants rated novel faces morphed with “positive” learned faces ($M=5.25$, $SD=.94$) more positively than faces morphed with “neutral” ($M=5.20$, $SD=.87$) or “negative” faces ($M=4.99$, $SD=.90$), $F(2, 106)=3.09$, $p<.05$, $\eta_p^2=.06$. There was also a significant main effect of morphing so that 35% morphs were evaluated more positively than 20% morphs, $F(1, 53)=11.14$, $p<.002$, $\eta_p^2=.17$. Importantly, these two main effects were qualified by a significant interaction, $F(2, 106)=3.81$, $p<.03$, $\eta_p^2=.07$, indicating that the learning generalization effect was more pronounced for the 35% than for the 25% morphs.

Although for the 20% morph condition, participants did evaluate morphs of positive faces more positively than morphs of negative faces (Figure 1.3A), the effect was not significant, $F_s<1$ 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 (Figure 1.3B), $F(1, 34)=3.46$, $p<.071$, $\eta_p^2=.09$. For the 35% morphs, the effect of behavior valence was significant (Figure 1.3C), $F(2, 106)=5.93$, $p<.004$, $\eta_p^2=.10$ 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 (Figure 1.3D), $F(1, 34)=10.48$, $p<.003$, $\eta_p^2=.24$ for the linear contrast.

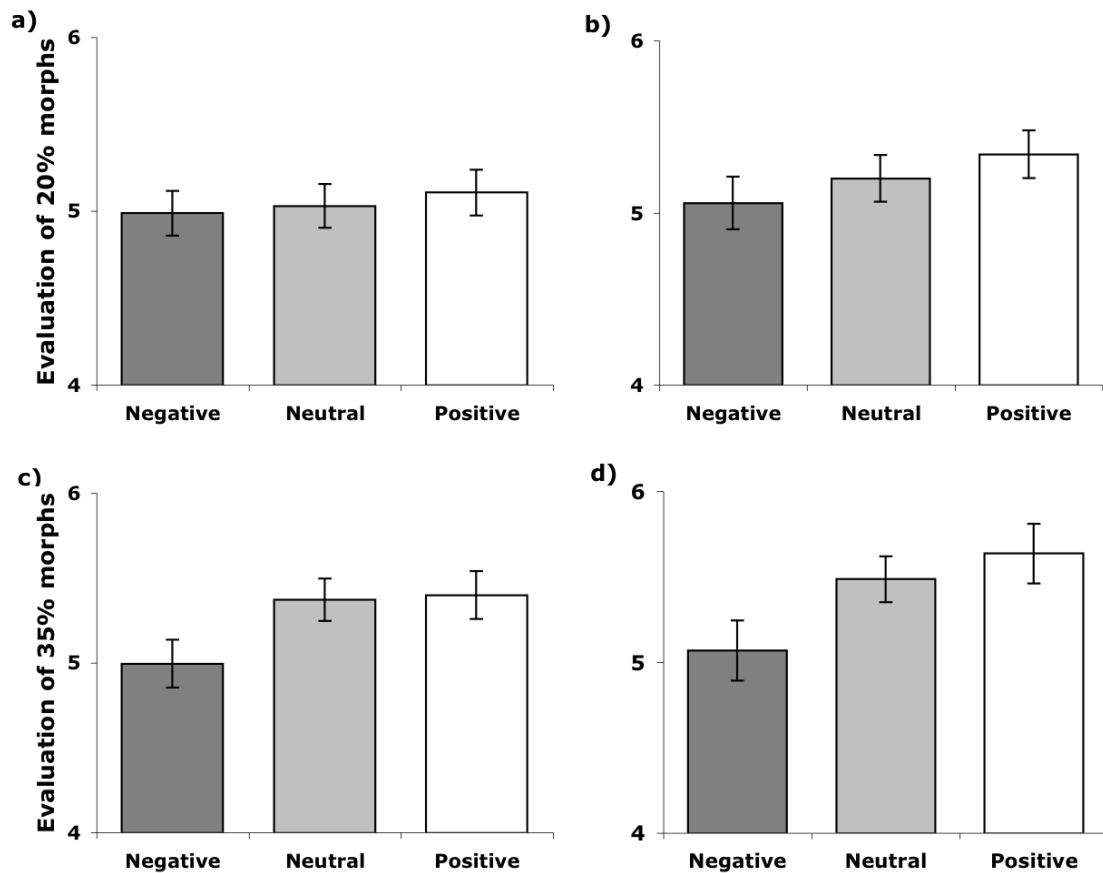


Figure 1.3. Behavioral results demonstrating learning generalization. (A) Mean trustworthiness ratings of the morphs containing 20% of the learned faces ($N=54$). Error bars in this graph and the following graphs represent standard error of the mean. (B) Mean trustworthiness ratings of the morphs containing 20% of the learned faces by participants who showed the expected relative ranking of learned face ratings ($N=35$). (C) Mean trustworthiness ratings of the morphs containing 35% of the learned faces ($N=54$). (D) Mean trustworthiness ratings of the morphs containing 35% of the learned faces by participants who showed the expected relative ranking of learned face ratings ($N=35$).

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 to 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. The experiments in the next part of the dissertation 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 (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 focuses on the familiarity of specific individuals. In this respect, our findings are related to Andersen and colleagues' findings that 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. 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.

Our account is also related to work suggesting that learning of incidental associations between facial features and personality characteristics could contribute to individual differences in face preferences (Hill, Lewicki, Czyzewska, & Schuller, 1990). For example, Hill et al. (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 to composites made up of faces paired with negative sounds. It is an interesting question to what extent the processes underlying the generalization of these kinds of associations differ from the generalization of associations based on person specific, attributional learning (e.g., Todorov & Uleman, 2003).

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 (looking) significant others, friends, and foes. This social face environment would most likely shape preferences for novel faces. Together, our findings suggest that general learning mechanisms based on similarity can account for idiosyncratic face preferences.

Finally, thinking back to the question posed in the introduction of the dissertation, our findings suggest that learning influences the perception of the learned faces. While learning has previously been found to change judgments based on facial appearance (Gross & Crofton, 1977; Todorov & Olson, 2008), it is not clear whether these changes

are due to changes in perception or to the influence of conceptual knowledge on evaluation. In the current study, we took a first step toward addressing this question by examining how learning influenced the evaluation of faces that resembled the learned faces – rather than the evaluation of the learned faces themselves. We found that learning generalized from the learned faces to the perceptually similar novel faces, suggesting that learning exerted its influence via changes in perception. The experiments in the following part of this dissertation further explore this idea by examining the mechanisms by which such learning generalization occurs.

Part 2: Mechanisms Underlying Affective Learning Generalization

(Adapted from Verosky & Todorov, submitted)

Physical similarity to familiar others can influence impressions based on facial appearance (DeBruine, 2002, 2005; Hill et al., 1990; Kraus & Chen, 2010; Lewicki, 1985; Verosky & Todorov, 2010; Zebrowitz, 1996; Zebrowitz et al., 2007; Zebrowitz et al., 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 the first part of this dissertation (henceforth referred to as Verosky & 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 learned faces. Participants were found to generalize their impressions of the learned faces to the similar-looking novel faces, such that faces similar to faces associated with negative behaviors were evaluated more negatively than those that were similar to faces associated with positive behaviors. While this study demonstrated that it is possible for social face environments to shape face preferences, it did not speak to the mechanisms underlying the effect.

In a series of three experiments, we used the paradigm developed in Verosky and Todorov (2010) to investigate the processes by which physical similarity influences the evaluation of faces. Experiment 2.1 examined whether learning generalization persisted even in the presence of relevant behavioral information about the morphed faces. Experiment 2.2 investigated whether participants need to be able to detect similarity between learned and morphed faces in order for similarity to influence face evaluation. Finally, Experiment 2.3 explored participants' control over their use of similarity information. While Experiment 2.1 speaks to the scope of learning generalization, Experiments 2.2 and 2.3 moved closer to directly examining the processes by which people use similarity information during face evaluation.

As a starting point, Experiment 2.1 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 et al., 2008; Carlston & Skowronski, 1994; 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. Thus, 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; Olson & Todorov, 2008; Rudoy & Paller, 2009). For example, Olson and Todorov (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.

While Experiment 2.1 investigated the use of physical similarity relative to use of behavioral information, Experiments 2.2 and 2.3 focused on similarity alone. In particular, Experiment 2.2 addressed the question of whether participants need to be able to detect similarity in order for it to influence evaluation. To give an example, say you met someone who resembled an old enemy of yours, but you were not able to see the

resemblance. Would you still distrust the person more than you would otherwise? Of course, to the extent that similarity influences evaluation, participants must be sensitive to this similarity, but the ability to detect similarity and the ability to use similarity information in evaluation need not have the same sensitivity.

Even if people need to be able to detect similarity between morphed faces and learned faces in order for learning generalization to occur, it is nevertheless possible that the use of similarity in judgments is a relatively automatic process not dependent on attentional resources and conscious recollection of the similarity. For instance, in Lewicki's (1985) study, participants were affected by the resemblance between the initial experimenter and the second potential experimenter, but none of them mentioned this factor when describing the reasons for their decision. Along these same lines, Hill and colleagues (1990) found that an association between face length and fairness in an original set of faces influenced participants' evaluation of subsequent faces, but none of them reported awareness of the manipulation. Finally, Andersen and Glassman (1999) found that subliminally presented words describing a significant other can trigger other features of the person that were not subliminally presented, demonstrating that transference can occur without intention.

Experiment 2.3 went on to examine the extent to which participants can control the influence of similarity information on 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 took a different approach. First, the experiment included a cognitive load manipulation. Cognitive load interferes with processes that are dependent on attention (Bargh, 1994; Todorov & Uleman, 2003). To the extent that

learning generalization persists in the presence of load, this would suggest that use of similarity information in face evaluation is characterized by, at least, one feature of automaticity: efficiency. Second, we manipulated the participants' intentions by instructing them to either include or exclude similarity information in their judgments. This is a validated procedure that allows for the estimation of controllability of cognitive processes (Jacoby, 1991; Payne, 2001, 2008). We compared learning generalization in the instructions conditions to learning generalization in the absence of instructions. To the extent that participants are able to increase and/or decrease their use of similarity information relative to baseline, this would imply controllable use of similarity information. We further examined the effect of load on instructions: to the extent that participants are able to control use of similarity information, we would expect for load to interfere with that control and to decrease the influence of the instructions on learning generalization.

Experiment 2.1

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. Then, they were asked to evaluate a series of morphs of novel faces with the learned faces. These morphs 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 learned faces would continue to influence evaluation of the morphed 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 morphed faces, we should see an effect of concurrent behavioral information on evaluation of the morphed faces, as well as an effect of learning generalization.

Method

Participants. Seventy-two participants recruited at a local shopping mall completed the task 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>). These photographs were selected from Kayser (1997). We chose three photographs of younger to middle-aged men to use as the learned faces (Figure 2.1A). 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.

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 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 of the task ended when participants answered both questions 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 of the three learned faces was morphed with 12 of the 36 novel faces (see Figure 2.1A for example stimuli). The faces were morphed to a level of 50%, using code from Steyvers (1999). While Verosky and Todorov (2010) used levels of morphing of 20% and 35%, a higher level of morphing was used in this and the following experiments in order to better examine the circumstances under which learning generalization

occurred or did not occur. Three of the morphs of each learned 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 et al., 1989). The pairing of faces and behaviors was counterbalanced across participants.

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 morphed 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 learned faces prior to the presentation of biographical information, learning in this and the subsequent experiments 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 by comparing pre- and post-ratings was largely redundant with learning assessed by comparing post-ratings alone.

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

Results

Learning. Sixty-nine percent of participants were able to identify the “good guy” and the “bad guy” after the first test round, 85% were able to do so after the second test

round, and 96% were able to do after the fourth test round. The biographical information influenced participants' evaluation of the learned 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, $t(70)=1.83$, $p=.07$.

Eleven of the 71 participants failed to evaluate the face paired with the negative biography more negatively than the face paired with the positive biography. Six of these 11 participants also required three or more test rounds to correctly identify the “good guy” and the “bad guy”. These 6 participants made up 60% of the group that required three or more test rounds to learn the associations between faces and biographies. Based on the convergence of these two learning measures, the data from these 11 participants (15% of the total number of participants) were excluded from further analysis.

Learning generalization. As expected, there was a main effect of concurrently presented behavior on participants' evaluation of the morphed faces (Figure 2.1B), $F(3, 177)=64.74$, $p<.001$, $\eta_p^2=.52$. Participants evaluated faces presented with negative behaviors ($M=3.95$, $SD=1.18$) more negatively than those presented with neutral ($M=5.30$, $SD=1.21$), $t(59)=6.84$, $p<.001$) or positive behaviors ($M=6.18$, $SD=1.17$; $t(59)=10.45$, $p<.001$), and those with neutral behaviors more negatively than those with positive behaviors, $t(59)=7.99$, $p<.001$. Participants also evaluated faces presented with negative behaviors more negatively than faces presented without behaviors ($M=5.23$,

$SD=1.07$; $t(59)=7.30$, $p<.001$) and those without behaviors more negatively than those with positive behaviors ($t(59)=7.19$, $p<.001$). They did not evaluate faces presented without behaviors differently from those presented with neutral behaviors ($t<1$).

Importantly, the effect of learning generalization was also significant (Figure 2.1B), $F(2, 118)=9.51$, $p<.001$, $\eta_p^2=.14$. Participants evaluated faces that were similar to the face paired with the negative biography ($M=4.91$, $SD=1.04$) more negatively than those that were similar to the face paired with the neutral ($M=5.23$, $SD=.90$; $t(59)=2.87$, $p=.006$) or positive biography ($M=5.36$, $SD=1.02$; $t(59)=3.91$, $p<.001$). 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, $t(59)=1.42$, $p=.16$. The interaction between concurrently presented behavior and learning generalization was not significant ($F<1$), indicating that learning generalization occurred regardless of concurrent behavioral information.

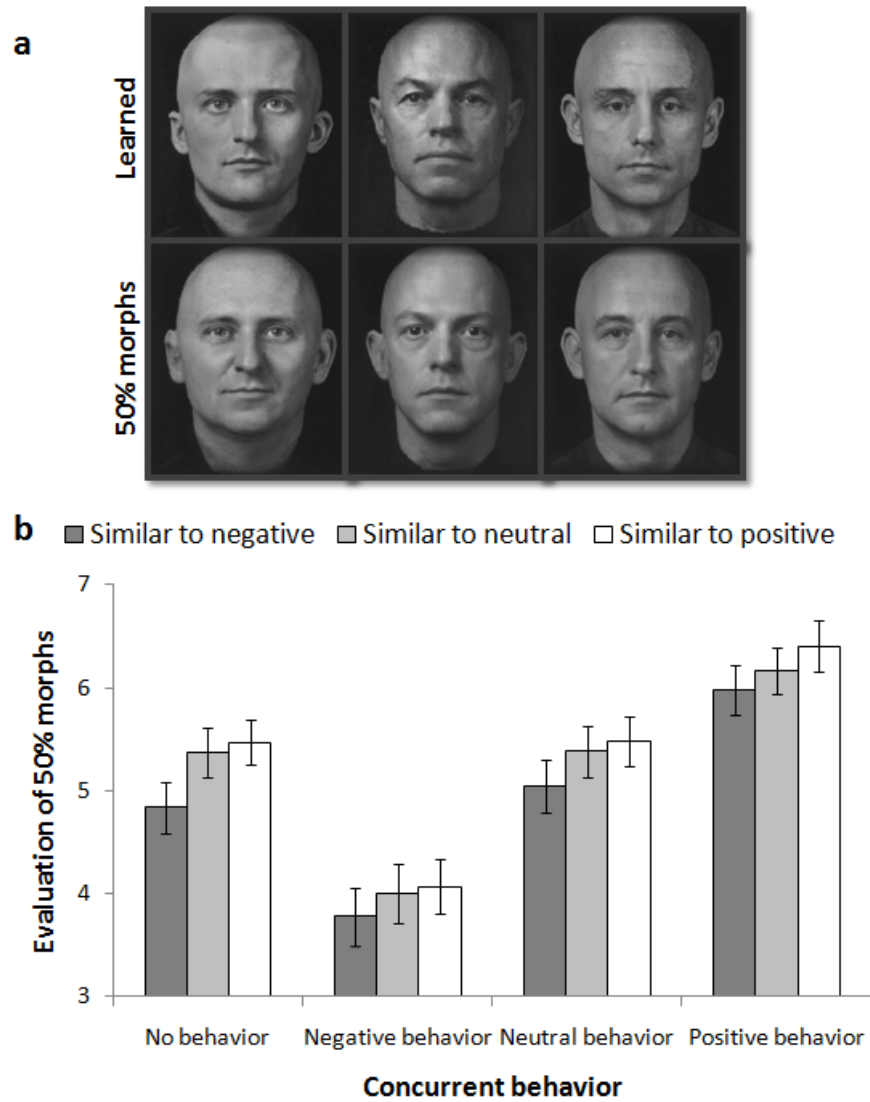


Figure 2.1. Learning generalization in the presence of concurrent behavioral information. (A) The three learned faces (top row) and examples of morphs containing 50% of those faces (bottom row). Morphs are shown directly below the faces they contain. (B) Mean trustworthiness ratings of morphs presented alone or presented with negative, neutral, or positive behaviors ($N=60$). The morphs contained 50% of the negative (dark gray), neutral (light gray), or positive (white) learned face. Error bars represent standard error of the mean.

Discussion

As expected, there was a large effect of concurrently presented behavioral information on evaluation of the morphed faces. Critically, there was also an independent effect of learning generalization. For example, morphed faces that were presented with negative behaviors were evaluated more negatively than those presented with no behaviors, neutral behaviors, or positive behaviors. However, within the group of morphed faces presented with negative behaviors, those faces that were similar to the negative learned face were evaluated more negatively than those that were similar to the positive learned 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.

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 learned 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.

Experiment 2.2

The second experiment was designed to investigate whether people need to be able to detect similarity in order for learning generalization to occur. In other words, if someone is not able to detect similarity between a learned face and a novel face, will that similarity still influence evaluation of the novel face? To test this, we had participants learn to associate faces with negative, neutral, and positive behaviors. Then, we asked participants to evaluate a series of morphed faces, interspersed with truly novel faces. Finally, we asked participants to categorize each of the faces they had previously seen as learned, similar to learned, or new. Since the same participants evaluated and categorized the faces, we were able to test whether learning generalization depends on participants' recognition of the similarity between the learned and the novel faces. In particular, we were interested in whether learning generalization would still occur for those morphed faces that were categorized as new.

Method

Participants. One hundred ten undergraduate students participated in this experiment for partial course credit.

Stimuli. Stimuli consisted of 87 black-and-white digital photographs of bald males. Fifty-four of the photographs were from the face set used in Experiment 2.1 (Goldstone & Steyvers, 1997) and the remaining 33 photographs were taken from Kayser (1997), the source of the original set of faces. Fifteen of these new faces were used in a distractor task (see below) and 18 were used as novel faces in the evaluation and categorization tasks.

Nine photographs of young to middle-age men were selected to be used as the learned faces. These nine faces were divided into three groups and participants learned to associate the groups with negative, neutral, or positive behaviors. The behaviors were from the database used in Experiment 2.1 (Fuhrman et al., 1989). The face and behavior groups were counterbalanced across participants.

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 2.1, 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.

Distractor task. After the learning task, participants completed a 2-back working memory task designed to interfere with their memory for the learned faces. In this task, participants viewed a series of faces and were instructed to press the spacebar whenever they saw a face that had appeared one face earlier (e.g., for the pattern ABA). Stimuli consisted of 15 faces and each face was shown five times, making a total of 75 trials. Twenty percent of trials were 2-back repetitions. Participants were presented with 7 practice trials to familiarize them with the procedure. Each trial began with a 500-millisecond fixation cross. The face remained on the screen for 2 s.

Navon task. We included a Navon task in this experiment to test the additional hypothesis that influencing face recognition would influence learning generalization. Navon letters are larger letters made up of repeating smaller letters (Navon, 1977), and reporting the identity of the smaller letters versus the large letter has previously been found to impair face recognition (Macrae & Lewis, 2002; Hills & Lewis, 2007). This effect is thought to result from a shift in processing orientation: reporting the identity of the smaller letters is thought to trigger a local processing orientation that is incompatible with the global orientation needed for face processing (Macrae & Lewis, 2002). Participants in the current study completed 80 trials of the Navon task after the distracter task and four trials prior to each trial of the evaluation and categorization tasks (see Hills

and Lewis, 2007, for a similar procedure). Participants were asked to report either the identity of the larger letter across all trials (global condition) or the smaller letters (local condition). The letters appeared in the center of the screen and participants chose one of four answer options at the bottom of the screen using the number keys 1-4. The global versus local condition did not influence recognition accuracy ($F < 1$ in all cases) and therefore we collapsed across this factor in subsequent analyses².

Face evaluation. Participants were asked to evaluate the trustworthiness of a series of faces. Participants were told that we were interested in first impressions and that there were no right or wrong answers. They were instructed to rely on their gut feeling and to work as quickly as possible. Each learned face was morphed with four novel faces, making a total of 36 morphs. Faces were morphed to a level of 50% using code from Steyvers (1999). To assess the ability to detect the similarity between the morphed faces and the learned faces, the 36 morphs were intermixed with 27 novel faces. Each trial began with a 500-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). The order of the morphs and novel faces was randomized for each participant. After evaluating the morphs and novel faces, participants were told that they would see the faces from the original task with the behaviors. Participants followed the same procedure to evaluate the nine learned faces.

Face categorization. During the final part of the experiment, participants were told that they would see a series of faces, including faces from the task with behaviors, faces that were similar to those learned faces, and new faces. Participants were instructed to categorize each face as learned (from the task with the behaviors), similar to learned,

or new. Each trial began with a 500-millisecond fixation cross. Then, a face was presented and remained on the screen until participants responded using the number keys 1(learned), 2(similar to learned), and 3(new). If participants categorized a face as learned or similar to learned, they were asked to further indicate whether the face itself or the face it resembled was previously shown with negative, neutral, or positive information. Participants responded to the second question using the number keys 4(negative), 5(neutral), and 6(positive). The order of the 9 learned, 36 morphed, and 27 novel faces was randomized for each participant.

Results

Learning. More than half of participants were able to correctly identify the valence of the behavior (negative, neutral, or positive) associated with each of the nine learned faces after two test rounds ($M=55\%$), eighty percent were able to do so after four test rounds ($M=81\%$), and ninety-five percent were able to do so after eight test rounds ($M=95\%$). After eight test rounds, the learning task ended.

The valence of the behavioral information influenced subsequent evaluation of the learned faces, $F(2, 206)=279.87, p<.001, \eta_p^2=.73$. Participants evaluated the faces paired with negative behavioral information ($M=2.51, SD=1.47$) more negatively than those paired with neutral ($M=5.04, SD=.86; t(103)=16.76, p<.001$) or positive information ($M=6.97, SD=1.70; t(103)=18.06, p<.001$). In turn, participants evaluated faces paired with neutral information more negatively than those paired with positive information, $t(103)=12.59, p<.001$. In total, data from 8% of participants were excluded from further analyses either because participants had not learned the face-valence associations after

eight test rounds (5%) or because they failed to rate the faces paired with negative behaviors more negatively than those paired with positive behaviors (3%).

Face categorization. Accurate categorization of faces as learned, similar to learned, or new depended on the type of face ($F(2, 198)=244.69, p<.001, \eta_p^2=.71$; Table 2.1). As expected, participants were more accurate at categorizing learned faces ($M=88\%$, $SD=14\%$) than novel faces ($M=82\%$; $SD=16\%$; $t(100)=3.14, p=.002$) or morphed faces ($M=42\%$, $SD=17\%$; $t(100)=21.83, p<.001$). In turn, participants were more accurate at categorizing novel faces as new than morphed faces as similar to learned, $t(100)=15.58, p<.001$.

After correctly categorizing learned faces as learned, participants were able to identify the valence of behavior that had previously been paired with the face most of the time ($M=86\%$, $SD=14\%$). After correctly categorizing the morphed faces as similar to learned, participants were able to identify the valence of behavior paired with the learned face that the morphed face resembled more than half of the time ($M=64\%$, $SD=17\%$).

Type of face	Categorization		
	“Learned”	“Similar to learned”	“New”
Learned	88 (14)	9 (10)	3 (9)
Morph	10 (9)	42 (17)	47 (18)
Novel	3 (6)	15 (13)	82 (16)

Table 2.1. Mean percent (and standard deviation) of the time learned faces, 50% morphs, and novel faces were categorized as “learned”, “similar to learned”, or “new”.

When participants did not categorize the morphed faces as similar to learned, they were more likely to categorize them as novel ($M=47\%$; $SD=18\%$) than as learned ($M=10\%$, $SD=9\%$; $t(100)=17.15$, $p<.001$). Meanwhile, when participants did not categorize the novel faces as new, they were more likely to categorize them as similar to learned ($M=15\%$, $SD=13\%$) than learned ($M=2.90$, $SD=5.68$; $t(100)=11.07$, $p<.001$).

Learning generalization to morphed faces. For the morphed faces, we were interested in whether the ability to detect similarity between the morphed faces and the learned faces would influence learning generalization. To examine this question, we compared learning generalization for trials where participants were able to correctly identify the valence of the learned face the morphed face resembled versus trials where participants categorized the morphed faces as new using a 2(detected similarity, did not) X 3(valence: negative, neutral, or positive) repeated measures ANOVA. We included only participants who had at least one observation per cell, which meant the analysis was based on 82% of participants and 74% of the available evaluation data.

As expected, there was a main effect of learning generalization ($F(2, 164)=9.43$, $p<.001$, $\eta_p^2=.10$), such that participants evaluated faces that were similar to negative learned faces ($M=4.85$, $SD=.93$) significantly more negatively than those that were similar to neutral ($M=5.07$, $SD=.97$; $t(82)=2.34$, $p=.02$) or positive learned faces ($M=5.25$, $SD=.97$; $t(82)=4.16$, $p<.001$). In turn, participants evaluated faces that were similar to neutral learned faces more negatively than those that were similar to positive learned faces, $t(82)=2.10$, $p=.04$. There was also a main effect of ability to detect similarity to the learned faces (Figure 2.2). Participants evaluated faces they later categorized as new ($M=4.91$, $SD=.91$) more negatively than those where they correctly

identified the valence of the similarity ($M=5.21$, $SD=.92$; $F(1, 82)=12.42$, $p=.001$, $\eta_p^2=.13$).

Importantly, these main effects were qualified by a significant interaction, $F(2, 164)=5.99$, $p=.003$, $\eta_p^2=.07$, indicating that learning generalization depended on participants' recognition of the similarity between the morphed faces and the learned faces. To elaborate, for the faces where participants were able to correctly identify the similarity (Figure 2.2A), there was a significant effect of learning generalization, $F(2, 164)=11.30$, $p<.001$, $\eta_p^2=.12$. As before, participants evaluated faces that were similar to negative learned faces ($M=4.82$, $SD=1.15$) significantly more negatively than those that were similar to neutral ($M=5.25$, $SD=1.31$; $t(82)=2.79$, $p=.007$) or positive learned faces ($M=5.56$, $SD=1.22$; $t(82)=4.56$, $p<.001$). Participants also evaluated faces that were similar to neutral faces more negatively than those that were similar to positive faces, $t(82)=2.04$, $p=.04$. However, for the faces categorized as new (Figure 2.2B), the effect of learning generalization was not significant, $F<1$. Participants did not rate the faces that were similar to negative learned faces ($M=4.88$, $SD=1.13$) significantly differently than those that were similar to neutral ($M=4.89$, $SD=1.02$) or positive learned faces ($M=4.95$, $SD=1.07$; t 's <1 for all pairwise comparisons).

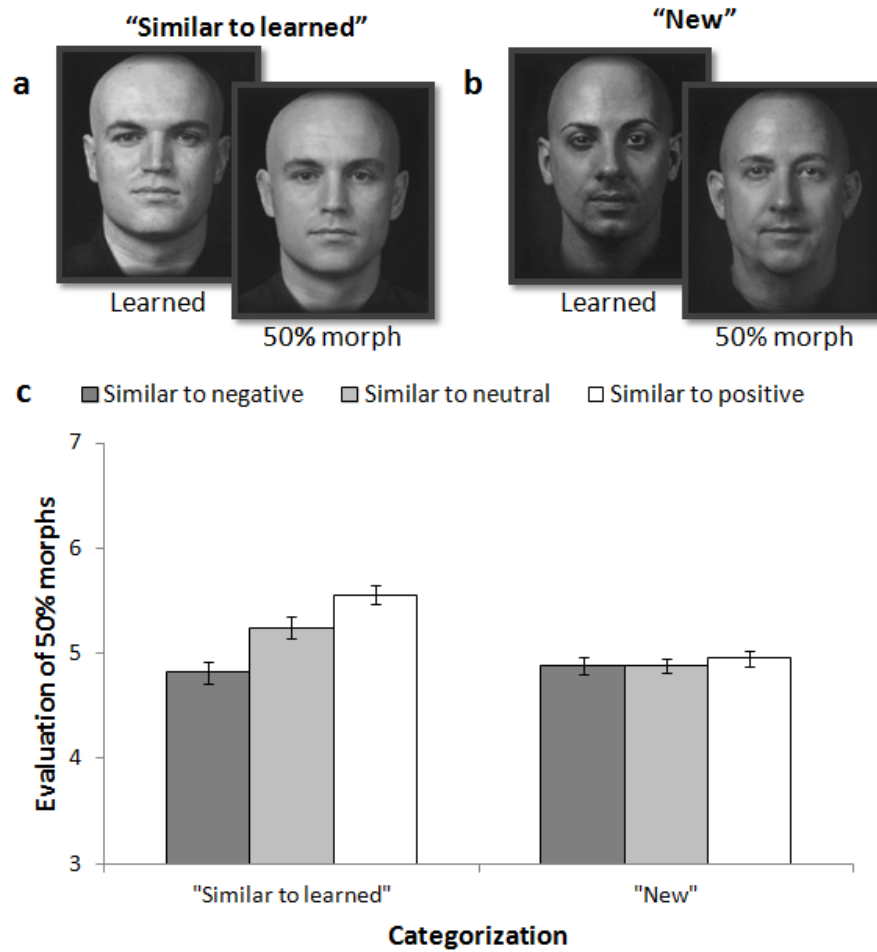


Figure 2.2. Learning generalization by recognition of physical similarity. (A) The morphed face (right) most frequently categorized as “similar to learned”. (B) The morphed face (right) most frequently categorized as “similar to new”. Each morph contains 50% of the learned face shown to its left. (C) Mean trustworthiness ratings of morphed faces where participants recognized the valence of the similarity versus faces categorized as new ($N=83$). The morphed faces contained 50% of the negative (dark gray), neutral (light gray), or positive (white) learned faces. Error bars represent standard error of the mean.

Learning generalization to novel faces. Although we were primarily interested in the evaluation of the morphed faces, we also examined whether the naturally occurring similarity between the novel faces and the learned faces influenced evaluation. As mentioned above, the novel faces were categorized as similar to the learned faces fifteen percent of the time ($M = 15\%$, $SD = 13\%$). To avoid dropping observations, we examined the influence of learning generalization on evaluation using a one-way ANOVA. The resulting analysis included data from 88% of participants.

There was a main effect of learning generalization, such that the type of learned face a truly novel face was seen as similar to mattered, $F(2, 182) = 10.56$, $p < .001$, $\eta_p^2 = .10$. Participants evaluated the faces that they saw as similar to negative learned faces ($M = 3.82$, $SD = 1.37$) significantly more negatively than those they saw as similar to neutral ($M = 4.70$, $SD = 1.56$; $F(1, 131) = 11.97$, $p = .001$, $\eta_p^2 = .08$) or positive learned faces ($M = 4.99$, $SD = 1.56$; $F(1, 120) = 19.38$, $p < .001$, $\eta_p^2 = .14$). In turn, participants evaluated the faces they saw as similar to neutral learned faces more negatively than those they saw as similar to positive learned faces, although this difference was not significant, $F < 1$.

Discussion

When participants detected the similarity between the morphed faces and the learned faces, categorizing the morphed faces as similar to learned, they evaluated those faces that were similar to negative learned faces more negatively than those that were similar to positive learned faces. However, when participants were not able to detect the similarity, categorizing the morphed faces as new, they did not evaluate faces that were similar to negative learned faces any differently than those that were similar to positive

learned faces. Thus, similarity only mattered when participants were able to detect it; when participants were not able to detect similarity, it did not influence evaluation.

At first, the finding that participants need to be able to detect similarity in order for learning generalization to occur may seem at odds with Verosky and Todorov (2010). In this previous study, learning generalization was observed for morphs containing 20% and 35% of learned faces, even though these morphs were not categorized differently than truly novel faces. However, different groups of participants performed the evaluation and categorization tasks in this study, and therefore it was not possible to examine whether learning generalization effects are contingent on recognition of similarity. There is a simple explanation for the apparent discrepancy: while the morphed faces and the truly novel faces were not treated differently, both were categorized as similar to the learned faces twenty percent of the time. In the current study, when we examined the effect of perceived similarity between novel faces and learned faces on evaluation, we found that those faces perceived as similar to negative learned faces were evaluated more negatively than those perceived as similar to positive learned faces. Thus, the categorization data in Verosky and Todorov (2010) should not be seen as indicating that participants were unable to detect similarity between the morphed faces and the learned faces; rather, they should be seen as demonstrating that participants were no more likely to detect the similarity between morphed and learned faces than the naturally occurring similarity between novel and learned faces.

In sum, this experiment indicates that participants need to be able to detect similarity between a morphed face and a learned face in order for this similarity to influence face evaluation. However, it is important to keep in mind that detection of

similarity need not be equated with controlled use of this information. We address this issue in the next experiment by orienting participants to similarity and asking them to use it in different ways.

Experiment 2.3

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). Within the domain of face evaluation, people’s impressions of others have been found to be sensitive to similarity along a number of dimensions, including age, emotion and health (Todorov et al., 2008; Zebrowitz, 1996; Zebrowitz & Montepare, 2008). Given the importance of physical similarity for a range of processes, we were interested in the possibility that use of similarity information in the evaluation of faces may be relatively automatic.

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. As in the first two experiments, participants in the current experiment learned to associate faces with behaviors and then evaluated a series of

morphs of novel faces with the learned faces. To examine control over use of similarity information, we introduced two manipulations during the evaluation of the morphed faces.

First, we introduced a cognitive load manipulation. 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). In the current study, we were interested in whether similarity would continue to influence face evaluation even when participants were under cognitive load.

Second, 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, 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 through use of instructions. Participants were either given no instructions regarding similarity, as in the previous experiments, or they were told that they would be seeing a series of faces that were similar to the learned 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). We examined control over use of similarity by comparing learning generalization in the instructions conditions to learning generalization in the baseline condition: If participants are able to increase their use of similarity information, we should see increased learning generalization in the inclusion condition relative to baseline; if participants are able to decrease their use of similarity information, we should see decreased learning generalization in the exclusion condition relative to baseline.

In addition to examining the main effect of instructions and load on learning generalization, we were further interested in how load would affect learning generalization under instructions. In particular, to the extent that participants are able to control their use of similarity information, we would expect that the effect of instructions would be diminished under load. For instance, if participants are able to increase their use of similarity information, learning generalization in the inclusion condition should decrease under load; on the other hand, if participants are able to decrease their use of

similarity information, learning generalization in the exclusion condition should increase under load.

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 2.1 (Goldstone & Steyvers, 1999). As in Experiment 2.2, nine photographs of young to middle-age men were selected to be used as the learned 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 the same as those used in Experiment 2.2. Once again, the face and behavior groups were counterbalanced across participants.

Procedures. *Learning.* Participants learned to associate faces with negative, neutral, and positive behaviors following the procedure from Experiment 2.2.

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 morphed face in the center of the screen, along with the learned and novel faces used to create it. This morph 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 morphed face and the learned 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 morphed face and the learned 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.

As in Experiment 2.2, each learned face was morphed with four novel faces, making a total of 36 morphs. Once again, 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.

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 2s. 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. The order of load and no load trials was randomized for each participant.

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

Results

Learning. Half of participants were able to identify the valence of behavior (negative, neutral, or positive) paired with each face after two rest rounds ($M=50\%$), seventy-five percent were able to do so after four test rounds ($M=75\%$), and nearly ninety-five percent were able to do so after eight test rounds ($M=93\%$).

There was a significant effect of the valence of the learned behavioral information on participants' evaluation of the learned 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 behavior significantly more negatively than those paired with positive behavior, $t(239)=15.70, p<.001$. In total, data from 12% of participants were excluded from further analyses either because participants had not learned the face-valence associations after eight test rounds (7%) or because they failed to rate the faces paired with negative behaviors more negatively than those paired with positive behaviors (6%).

Accuracy of cognitive load task. Participants were able to correctly indicate whether a number was old or new 91% of the time ($SD=10\%$). Accuracy was significantly higher than chance (chance=50%; $t(224)=61.87, p<.001$), but significantly lower than perfect performance ($t(224)=13.06, p<.001$). Accuracy did not significantly differ by instruction condition, $F(2,222)=1.26, p=.29, \eta_p^2=.01$.

Learning generalization. There was a significant main effect of learning generalization (Figure 2.3; $F(2, 444)=109.14$, $p<.001$, $\eta_p^2=.33$), with participants evaluating faces that were similar to negative learned faces ($M=4.66$, $SD=.89$) more negatively than those that were similar to neutral ($M=5.06$, $SD=.81$; $t(224)=7.42$, $p<.001$) or positive learned faces ($M=5.44$, $SD=.90$; $t(224)=10.62$, $p<.001$). In turn, participants evaluated neutral learned faces more negatively than positive learned faces ($t(224)=6.92$, $p<.001$).

There was also a main effect of load ($F(1, 222)=9.91$, $p=.002$, $\eta_p^2=.04$), with participants evaluating faces more positively when they were under load ($M=5.10$, $SD=.72$) versus not under load ($M=5.01$, $SD=.72$). This main effect was qualified by a significant interaction with learning generalization ($F(2, 444)=5.47$, $p=.005$, $\eta_p^2=.02$). Although learning generalization was significant regardless of whether participants were under load, it was stronger when participants were not under load ($F(2, 448)=71.76$, $p<.001$, $\eta_p^2=.24$) versus under load ($F(2, 448)=47.26$, $p<.001$, $\eta_p^2=.17$). In both cases, all pairwise comparisons between evaluation of faces that were similar to negative, neutral, and positive morphed faces were significant ($t(224) \geq 3.76$, $p<.001$; see Table 2.2 for means).

	Similar to negative	Similar to neutral	Similar to positive
No load	4.60 (.95)	4.96 (.89)	5.46 (1.03)
Load	4.72 (.98)	5.17 (.90)	5.42 (1.00)

Table 2.2. Mean (and standard deviation) trustworthiness ratings of morphed faces containing 50% of negative, neutral, or positive learned faces when participants were under cognitive load versus no load.

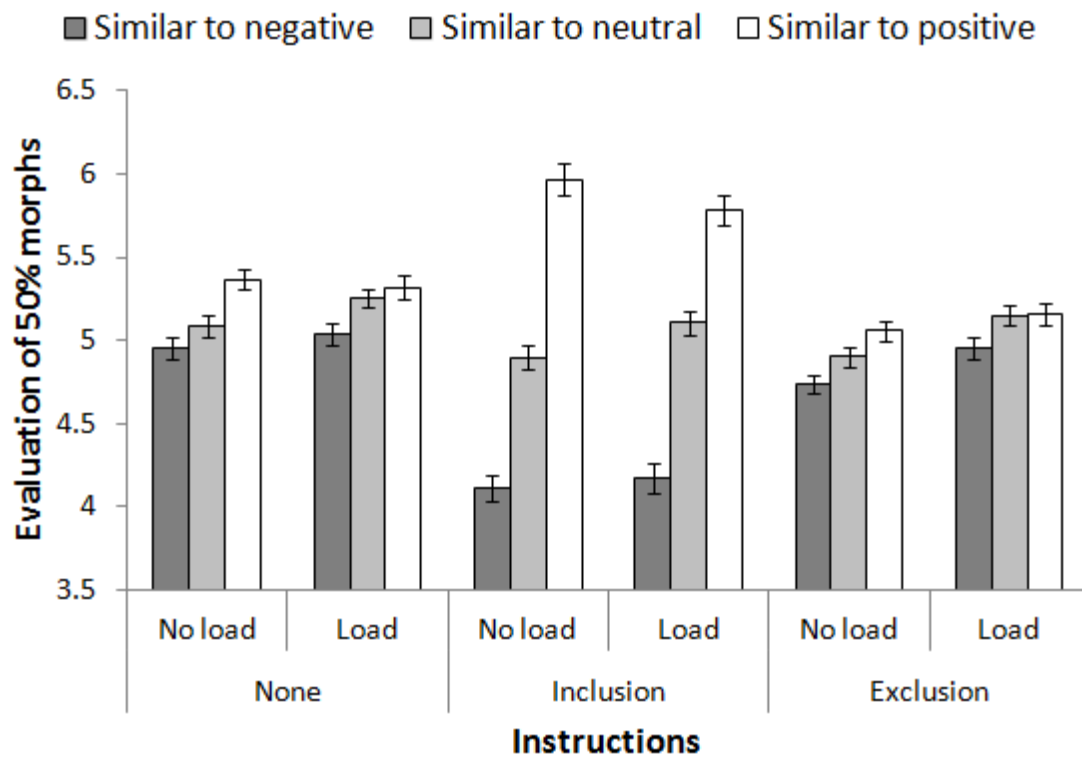


Figure 2.3. Use of similarity information in learning generalization. Mean trustworthiness ratings of morphed faces as a function of instructions and cognitive load ($N=225$). Participants were not told about the similarity between the morphed faces and the learned faces (baseline), or they were asked to either take similarity into account (inclusion) or to avoid taking it into account (exclusion). The morphed faces contained 50% of the negative (dark gray), neutral (light gray), or positive (white) learned faces. Error bars represent standard error of the mean.

There was no main effect of instruction condition on face evaluation ($F(2, 222)=1.55, p=.21, \eta_p^2=.01$), but there was a significant interaction between instructions and the strength of learning generalization ($F(4, 444)=40.49, p<.001, \eta_p^2=.27$). In particular, there was a stronger effect of learning generalization in the inclusion condition as compared to the baseline ($F(2, 296)=49.61, p<.001, \eta_p^2=.25$) and exclusion conditions ($F(2, 296)=58.64, p<.001, \eta_p^2=.28$), indicating that participants were able to increase their use of similarity information when asked to do so. On the other hand, the effect of learning generalization 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. Further exploring the interaction between the inclusion condition versus the baseline and exclusion conditions, participants in the inclusion condition evaluated faces that were similar to negative learned faces ($M=4.14, SD=.85$) significantly more negatively and faces that were similar to positive learned faces ($M=5.87, SD=.90$) significantly more positively than participants in the other two conditions ($M=4.92, SD=.79; t(223)=6.78, p<.001$; and $M=5.22, SD=.83; t(223)=5.38, p<.001$, respectively). Participants did not evaluate faces that were similar to neutral learned faces ($M=5.00, SD=.80$) significantly differently across conditions ($M=5.10, SD=.82; t<1$).

Finally, the interaction between load and instruction condition was not significant, $F(2, 222)=2.70, p=.07, \eta_p^2=.02$, nor was the three-way interaction between load, valence, and instruction condition, $F<1$. However, planned comparisons revealed that the effect of load on learning generalization was more pronounced in the inclusion condition ($F(2,$

148)=3.27, $p=.04$, $\eta_p^2=.04$) versus the baseline ($F(2, 148)=1.48$, $p=.23$, $\eta_p^2=.02$) and exclusion conditions ($F<1$).

Discussion

As in the previous experiments, participants evaluated faces that were similar to negative learned faces more negatively than faces that were similar to positive learned faces. Importantly, a cognitive load manipulation only slightly diminished this learning generalization effect, suggesting an efficient use of similarity information. On the other hand, instructions regarding use of similarity information had a larger impact on learning generalization, suggesting that the use of this information was controllable.

Specifically, instructions to take similarity into account led to a stronger effect of learning generalization in the inclusion condition than in the baseline and exclusion conditions, indicating that participants were able to increase their use of similarity information when asked to do so. However, at the same time, instructions to avoid taking similarity into account did not eliminate the learning generalization effect entirely, indicating that participants were not able to completely disregard 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.

As noted above, 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 negative learned faces more negatively than those that were similar to positive learned faces. 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.

General Discussion

Across three experiments, participants evaluated faces that were similar to negative learned faces more negatively than those that were similar to positive learned faces. In Experiment 2.1, learning generalization persisted even in the presence of relevant behavioral information. In Experiment 2.2, learning generalization was found to occur when participants were able to detect similarity between morphed faces and learned faces, but not when they were unable to detect similarity. Finally in Experiment 2.3, 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. Thus, learning generalization persisted under a variety of circumstances, indicating that use of similarity is a powerful process.

Together, Experiments 2.2 and 2.3 paint a nuanced picture of use of similarity information in learning generalization. While Experiment 2.2 demonstrates that participants need to be able to detect similarity in order for learning generalization to occur, Experiment 2.3 cautions against equating detection of similarity with controlled use of this information. In Experiment 2.2, we did not ask participants about similarity until after they had evaluated all of the morphed faces. While this ensured that participants' use of similarity information was not influenced by the task of similarity detection, it also meant that the results did not speak to how participants used the similarity information as they were evaluating the faces. To address this question, Experiment 2.3 went on to explore participants' control over their use of similarity information. We found that participants continued to use similarity information even when they were under cognitive load. In addition, we found that while participants were able to increase their use of similarity information, they were not able to exclude similarity information from judgments altogether. Cognitive load influenced learning generalization in the inclusion condition but not in the other conditions, suggesting that participants were able to control increased use of similarity information but not use of similarity in the first place. Thus, while people need to be able to detect similarity in order for it to influence face evaluation, use of similarity information appears to be relatively automatic.

Although we did not directly assess whether participants intended to use similarity information, Experiments 2.1 and 2.3 address this issue, albeit indirectly. In Experiment 2.1, 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 2.3, 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.

Our findings also speak to the broad scope of learning generalization. In particular, the persistence of learning generalization in Experiment 2.1 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, the fact that learning generalization occurred for truly novel faces that were perceived as similar to the learned faces in Experiment 2.2 demonstrates that naturally occurring similarity (even that observed within the confines of an experiment) can trigger learning generalization. Finally, participants' inability to completely exclude similarity information from their judgments in Experiment 2.3 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; b) when similarity has not been manipulated; and c) despite intentions to ignore it.

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

an original set of learned faces can influence evaluation of a second set of morphed faces (Verosky & Todorov, 2010). 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; Zebrowitz et al., 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 varying the circumstances under which learning generalization occurs in order to examine the processes underlying use of physical similarity in face evaluation. While there may well be differences in the exact mechanisms underlying use of similarity in the cases described above, given the rapid learning and subtle similarity in the current studies, it seems likely that many of the findings will apply to these other cases as well.

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 2.1 demonstrated that people continue to use physical similarity in their judgments of faces, even when there is behavioral information available. Experiment 2.2 showed that participants need to be able to detect the similarity between faces in order for it to influence evaluation. Finally, Experiment 2.3 showed that while participants are able to increase their use of similarity information, they are not able to disregard it. Together, these findings show that learning generalization is a powerful and relatively automatic process, which influences the evaluation of faces under a variety of circumstances.

Part 3: Representations of Individuals in Ventral Temporal Cortex Defined by Faces and Biographies

(Adapted from Verosky, Todorov, & Turk-Browne, submitted)

Multiple regions in ventral temporal cortex, including the lateral fusiform gyrus (FG), respond more strongly to faces than to other categories of objects (Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997; Tsao, Moeller, & Freiwald, 2008). However, since individual faces convey both category-level and exemplar-level information, these regions could be involved in the categorical detection of faces or the differentiation between specific face exemplars. Neuroimaging techniques such as fMRI adaptation and multi-voxel pattern analysis (MVPA) have been used to address the question of what these regions represent, with mixed results.

Univariate analyses of the FG have revealed adaptation for repeated versus novel facial identities (Rotshtein, Henson, Treves, Driver, & Dolan, 2005; Verosky & Turk-Browne, 2012). In such studies, the FG response measured by functional magnetic resonance imaging (fMRI) is attenuated when two faces depicting the same identity are shown sequentially, despite changes in physical properties such as viewpoint.

Multivariate analyses have produced less conclusive evidence. Although distributed patterns of activation in ventral temporal cortex convey information about object categories (Haxby et al., 2001), patterns associated with specific identities have been found less reliably (Kriegeskorte, Formisano, Sorger, & Goebel, 2007; Nestor, Plaut, & Behrmann, 2011), in contrast to the anterior temporal lobe (ATL).

While previous studies have tended to define identity in terms of visual features, identity representations in everyday life are much more complex. For instance, mere

acquaintances see each other multiple times, know each other's names, and often even know detailed social information about each other (e.g., group membership). Becoming familiar with someone in one or more of these ways alters face processing in the brain: familiar faces engage a broad network of regions more strongly than unfamiliar faces (Gobbini & Haxby, 2006, 2007; Todorov, Gobbini, Evans, & Haxby, 2007; Cloutier, Kelley, & Heatherton, 2011). Many of these regions are tied to higher-level emotion and social cognition, and thus we sought to explore how visual and social familiarity influences identity representations in perceptual brain regions.

Behaviorally, face recognition is superior for familiar versus unfamiliar faces across changes in visual properties (such as lighting, viewpoint, context, and emotional expression; Johnston & Edmonds, 2009), indicating that experience produces more invariant facial identity representations. Research in other domains further supports this notion: In the developmental literature, for example, teaching infants to individuate versus categorize monkey faces with verbal labels prevents a later decline in the ability to discriminate between those faces resulting from perceptual narrowing (Scott & Monesson, 2009). In the social psychological literature, warning participants about the cross-race recognition deficit — that face recognition is worse for faces from other races — can reduce or eliminate the effect, suggesting that the deficit partly results from categorization of out-group members (Hugenberg, Miller, & Claypool, 2007). Together, these studies suggest that learning to individuate others through everyday interactions, verbal labels, or by increasing motivation, improves face processing.

In the current study, we explored whether experience and training with a set of identities would result in richer identity representations that could be decoded more

robustly from ventral temporal cortex. Prior to scanning, participants completed a behavioral training task in which they saw multiple identities from different viewpoints and learned to associate each identity with a unique name. For half of the identities, participants also learned fictional biographical information about that person. During a subsequent fMRI session, participants were presented with sequential pairs of faces followed by pairs of names. For each pair of faces, the two faces depicted either the same identity or different identities. In univariate analyses, we tested for voxels that showed an attenuated response to pairs of faces where the identity was the same versus different (fMRI adaptation or repetition suppression; Grill-Spector, Henson, & Martin, 2006; Turk-Browne, Scholl, & Chun, 2008). In multivariate analyses, we created an activation map for each identity and correlated these maps across runs within different regions of interest (ROIs). We tested whether the activation maps for a given identity were more highly correlated with themselves than with those for other identities, which would provide evidence of distributed identity representations. We assessed such representations in ventral temporal cortex, where current evidence for them is mixed, and in the anterior temporal lobe, where they have been observed more reliably (Kriegeskorte et al., 2007; Nestor et al., 2011).

Additionally, we tested how learning biographical information influences identity representations. We considered two possible ways that this could occur: First, biographical information might enhance the strength of identity representations, in which case the correlation of an identity with itself should be higher for identities associated with biographical information than those that were not. Second, biographical information might alter the nature of identity representations by adding a new dimension of

information, in which case identities might be more correlated with other identities with the same amount of biographical information.

Methods

Participants

Twenty-two volunteers participated in the study for monetary compensation (5 males; mean age=22.2 years, range=18-37 years). Informed consent was obtained for a protocol approved by the Princeton University Institutional Review Board. All participants were right-handed, had normal or corrected-to-normal vision, and reported no history of neurological or psychological disorders. The data from two participants were excluded from analysis due to excessive head motion (greater than 3 mm in one or more runs), resulting in a final sample of 20 participants.

Stimuli

Color photographs of 16 female facial identities from the Karolinska Directed Emotional Faces set (Lundqvist, Flykt, & Ohman, 1998) were used as stimuli. For each identity, there were three photographs: one facing forward and two with the head turned 30-40° to the left and right. The individuals displayed neutral facial expressions and eye gaze matched head direction.



Figure 3.1. Tasks. (A) Training task. On each ‘learning’ trial, participants were shown a name followed by three views of that facial identity. For faces associated with biographies, the biographical information appeared immediately after the name but before the three views of the face. On each ‘test’ trial, participants were shown views of three different facial identities followed by a name and they chose which face matched the name. (B) Main task. On each trial, participants were shown a sequence of two faces followed by two names. Participants chose which name matched the second face. (C) The first and second face in the main task depicted either the same person (identity-same) or different people (identity-change).

Procedure

Training task. Either the day before the scan session or earlier in the day of the scan session, participants completed a training task in which they learned to associate faces with names (Figure 3.1A). Each of the 16 faces was randomly assigned to a unique name selected from the list of the twenty most popular female baby names for 1989 (<http://www.ssa.gov/oact/babynames/>); this year was chosen to ensure that our participants would be very familiar with the names (i.e., it matched their estimated birth

year). In addition, eight of the 16 faces were randomly paired with short biographical vignettes (mean words =186, SD=9). Four vignettes were primarily negative in tone and four were primarily positive (see Appendix B). We were not interested in valence effects *per se* due to limited power for examining such differences (four identities each for positive and negative), but rather used valenced information to make the vignettes memorable.

The training task consisted of two phases, a learning phase and a test phase, that participants cycled between until reaching a criterion. During the learning phase, participants saw 16 identities, presented one at a time in a random order. For each identity, a name appeared in the center of the screen. When the participant pressed the space bar, the task forwarded to the next screen. For identities paired with biographical information, a paragraph of the biographical vignette appeared in the center of the screen. Each vignette was broken into two paragraphs, and these two paragraphs were shown on alternating blocks of the training task. Participants were instructed to read the biographical information carefully and to try to form an impression of the person based on it (Hamilton, Katz, & Leirer, 1980). While participants were not asked to recall the biographical information during the training or main tasks, they were given a surprise memory test at the end of the study (see Supplementary Methods and Results). When the participant pressed the space bar again, a sequence of three images appeared in the center of the screen. This screen immediately followed the name screen for identities that were not associated with biographical information. The first image showed the person facing to the left, the next image facing forward, and the last facing right. Each image appeared for 5 s, with the name underneath. Because the biographical information was presented prior

to the faces, participants viewed the faces with and without biographical information for an equal amount of time.

During the test phase, participants were tested on which name belonged to each face. On each trial, a sequence of three faces appeared in the center of the screen. Each face was shown for 5 s, and after the last face, a name appeared. Participants indicated the serial position of the face that belonged with the name using the number keys 1-3. Participants received feedback about whether their response was correct. Each test block consisted of 48 trials, such that every view of each face was the correct answer for one trial. After each test block, participants received an accuracy score. To ensure that participants saw both paragraphs of each vignette, they were required to complete at least two cycles of the task. After the second test block, the learning-test cycles ended when participants achieved greater than 95% accuracy in the last test block. Immediately before the scan, participants completed an additional cycle of the learning phase of the training task, where they were presented with the entire vignette for each identity paired with biographical information.

Main task. During the scan session, participants viewed stimuli on a projection screen at the back of the scanner bore via an angled mirror attached to the head coil. On each trial, participants were presented with a sequence of two faces in the center of the screen (Figure 3.1B). The faces subtended $5.0 \times 6.8^\circ$, and appeared for 300 ms each separated by a 400 ms inter-stimulus interval. Two names appeared immediately afterwards, located above and below fixation. One name belonged to the second face and the other was a foil that did not belong to either face (vertical position randomized). Participants were instructed that the identity of the two faces in each pair could either

remain the same or change, but regardless, their task was to indicate which name belonged to the second face. They pressed a button with their right index finger to select the name above fixation and with their right middle finger for the name below fixation. The two names appeared for 2.5 s, regardless of when participants responded. There was a randomly selected jittered inter-trial interval of 1.0 or 2.5 s, resulting in trial onset asynchronies of 4.5 or 6.0 s. To ensure familiarity with this task, participants completed practice runs at the end of the behavioral training session until they reached a criterion of greater than 90% correct on a single run.

The fMRI response for each identity was assessed using *identity-same* trials, where a particular identity was repeated. To increase the generalizability of findings across changes in stimuli, we included two types of runs: view-repeat runs, where the viewpoint of the faces in each pair was the same, and view-change runs, where there was a change in viewpoint from the first to the second face. Both run types had a 2 (social knowledge factor: name-and-biography, name-only) x 2 (identity repetition factor: identity-same, identity-change) design. For the view-repeat runs, there were three viewpoint conditions: the two faces in each trial faced to the left (left-left), right (right-right), or forward (center-center). For the view-change runs, the changes in view were limited to less than 45°, resulting in four conditions: left-center, right-center, center-left, and center-right. To match the frequencies of viewpoints in the view-repeat runs to the view-change runs, there were twice as many center-center trials as left-left and right-right trials. For both view-change and view-repeat runs, there were an equal number of the identity-same and identity-change trials, and the frequency of each viewpoint and identity was counterbalanced within each of these conditions. Participants completed eight

functional runs, alternating between view-repeat and view-change runs (starting run type counterbalanced across participants). Each run contained 64 trials and lasted 364.5 s. Collapsing across viewpoint, each identity appeared in the identity-same condition 16 times.

fMRI Acquisition

Data were collected on a Siemens 3T Allegra head-only scanner using a Nova Medical head coil. Functional images were acquired using a T2*-weighted echo planar imaging (EPI) sequence, with a 1500 ms TR, 28 ms TE, 64° flip angle, and a 64x64 matrix. Twenty-six interleaved oblique axial slices parallel to the anterior commissure/posterior commissure line were used to achieve whole-brain coverage (3.5x3.5 mm inplane, 5 mm thick). A T1-weighted high-resolution anatomical image was acquired with an MPRAGE sequence for registration purposes. In all but three participants, a second T1-weighted image coplanar with the functional slices was acquired to assist with registration.

fMRI Analysis

Preprocessing. Analyses were conducted with FEAT in FSL (www.fmrib.ox.ac.uk/fsl). The first six volumes of each functional run were discarded to allow for T1 equilibration. Data in each run were corrected for slice-acquisition time and head motion, spatially smoothed (5 mm FWHM), de-trended, and high-pass filtered (128 s period cutoff). The data from two of four runs of the same view-repeat/view-change run type were then concatenated, resulting in four concatenated runs total (two view-repeat runs and two view-change runs). This was done because the frequency of identities was counterbalanced across pairs of runs. The concatenated functional runs were registered to

the coplanar anatomical scan and then to the high-resolution anatomical scan. For the participants without a coplanar anatomical scan, the functional runs were registered directly to the high-resolution anatomical scan. Data were normalized into Montreal Neurological Institute (MNI) space and interpolated to 2 mm isotropic voxels.

General linear modeling. Because of the rapid presentation of the two images in each trial and the sluggishness of the blood oxygen level-dependent (BOLD) response, we examined the combined response to pairs of images (e.g., Verosky & Turk-Browne, 2012). There were four identity-same trials for each identity in each concatenated run, which were used to estimate the voxelwise response to each identity in a general linear model (GLM). The model included a separate regressor for each of the 16 identities, including only identity-same trials. We also included a regressor for missed and incorrect responses and two regressors for the identity-change condition: one for pairs of name-and-biography identities and the other for pairs of name-only identities (trials with different identities always contained two identities from the same social knowledge condition). Regressors were convolved with a double-gamma canonical hemodynamic response function. Temporal derivatives were included for each regressor to account for variability in the latency of the BOLD response. Motion correction parameters were included as covariates of no interest.

Identity adaptation. fMRI adaptation was assessed by comparing BOLD responses for identity-change versus identity-same trials (e.g., Rotshtein et al., 2005; Verosky & Turk-Browne, 2012; Figure 3.1C). This contrast was performed separately for name-and-biography and name-only identities. Furthermore, to assess whether social knowledge influences the strength of identity adaptation, we tested whether there was

greater adaptation for name-and-biography identities than for name-only identities. For group analyses, we used *randomise* in FSL to perform non-parametric permutation t-tests for the contrasts of name-and-biography adaptation, name-only adaptation, and name-and-biography > name-only adaptation. The resulting maps were thresholded at $p < .05$ (cluster-corrected), with a cluster-forming cutoff of $p < .001$.

Pattern analyses. The analyses above examined identity information within individual voxels. To test for identity information distributed in patterns of activation across voxels, we created a voxel map for each identity and correlated these maps across runs. For each participant, we created these maps by contrasting activation for all identity-same trials of a given identity with the activation for all identity-change trials. Although the identity-change baseline was identical for all faces (within the same social knowledge group) in a given run, subtracting this baseline allowed us to remove activation reflecting the mere presence of faces. This functioned to mean-center the data, such that the voxel map for each identity reflects whether a voxel responded more or less than average. The resulting correlations between maps thus reflect the *change* in pattern correlation for a given identity. The map for each identity was correlated with the maps for all other identities in the same social knowledge group, producing one 8×8 correlation matrix for the name-and-biography identities and another one for the name-only identities (Figure 3.2A, 3.2B). Correlations were computed separately for view-repeat and view-change runs, resulting in four total matrices.

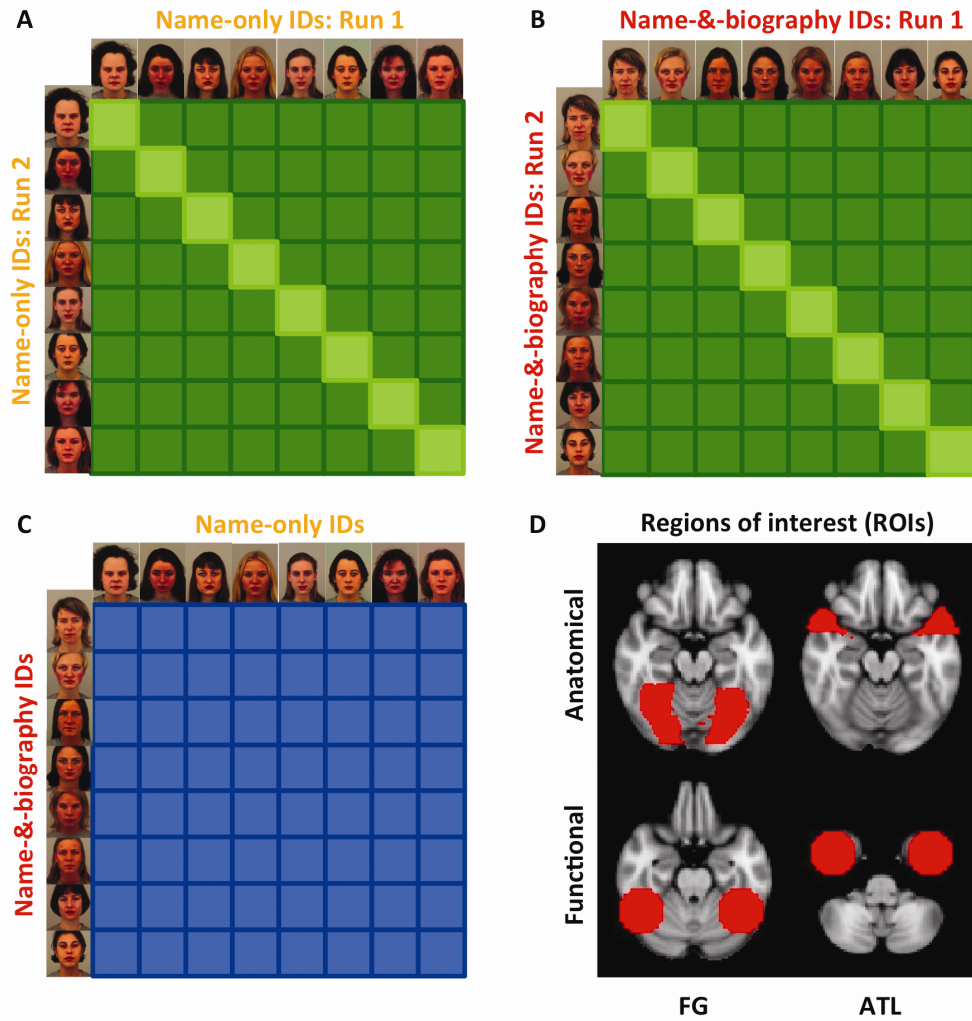


Figure 3.2. Pattern analyses. An activation map was created for each identity and these activation maps were correlated across runs. To test for identity-specific representations, the correlations were calculated for name-only identities (A) and name-and-biography identities (B) and the resulting within-identity correlations (light green) were compared to within-group correlations (dark green). To test for group-level representations, the activation maps for name-only identities were correlated with name-and-biography identities across runs (C) and the resulting between-group correlations (blue) were compared to the within-group correlations (dark green). Pattern analyses were conducted over maps masked with one of eight ROIs (D).

For each matrix, correlations along the diagonal reflected the correlation of each identity with itself across runs (within-identity), while the correlations off the diagonal reflected the correlations for different identities in the same social knowledge group across runs (within-group). If specific identities have unique voxel patterns, then the within-identity correlations should be greater than within-group correlations. We tested this hypothesis with a permutation test: For each participant, we shuffled the columns of each correlation matrix in every possible way ($8!=40,320$) and calculated the mean of the diagonal in each permuted matrix. We then combined data across social knowledge groups and across the view-repeat and view-change runs by taking the mean of the four diagonal means for each permutation, and then across participants by calculating the group mean. Finally, we computed the probability of the true observed group mean by dividing the number of permutations with means higher than the true mean by the total number of permutations. Probabilities less than .05 indicate the presence of reliable activation patterns for specific identities. We examined correlations separately for name-and-biography and name-only identities below, but combined across view-repeat and view-change runs for all permutation tests to increase the statistical power of the tests, as well as to increase the generalizability of the findings across stimulus differences.

Creating separate correlation matrices for name-and-biography and name-only identities allowed us to examine identity representations separately for each social knowledge group. To test whether learning biographical information produced more unique identity representations, we compared the strength of within-identity correlations between the two groups using a similar approach to above: For each participant, we shuffled the columns of each correlation matrix in every possible way. At each

permutation, we calculated the mean of the diagonal (within-identity) correlations for the name-and-biography and name-only groups, and then subtracted the latter from the former. We then combined data across participants by calculating the mean of this difference. Finally, we computed the probability of the true mean difference by dividing the number of permutations with values higher than the true difference by the total number of permutations. We had a slightly directional hypothesis that social knowledge would increase the strength of identity correlations, but we also wanted to consider the possibility that learning social knowledge might decrease the need for reliance on visual representations and thus reduce correlations. Therefore, we conducted a two-tailed test: $p < .025$ means that identity correlations were stronger for name-and-biography vs. name-only identities, while $p > .975$ means the reverse.

In addition to examining patterns for specific identities, we also examined whether learning biographical information changed the nature of identity representations in a more general manner. Specifically, biographical details may serve as a feature in the neural representation of identity. The presence of information along this feature dimension for all name-and-biography identities and not for name-only identities may have resulted in higher correlations within each social knowledge group than between groups (Figure 3.2C). Such a grouping effect would suggest that biographical information had been coded in a region, but in a categorical manner that did not contribute to the decoding of specific identities.

To examine group-level correlations, we tested whether the mean correlation between different identities from the same group (within-group) was higher than the mean correlation between identities from different groups (between-group). For each

participant, the 16 identity maps from one run were correlated with those from the other run of the same type (i.e., view-repeat or view-change), resulting in two 16×16 matrices. Shuffling the matrix columns in every possible way would result in a huge number of permutations, and therefore the columns were shuffled 40,320 times (to match the identity analyses). For each permutation, we calculated the mean within-group correlation (i.e., name-and-biography with other name-and-biography, name-only with other name-only) and between-group correlation (i.e., name-and-biography with name-only), based on the original column labels. We then subtracted the between-group correlation from the within-group correlation and combined data across participants by calculating the mean difference for each permutation. Finally, we computed the probability of the true mean difference by dividing the number of permutations with values higher than the true mean by the total number of permutations. Probabilities less than .05 indicated that the patterns contained information about whether an identity had been linked to social knowledge.

Regions of interest. Pattern analyses were conducted for ROIs in the bilateral FG and bilateral ATL (Figure 3.2D). To ensure that appropriate regions were included in the analysis, and to attempt to replicate prior studies, ROIs were defined both anatomically and functionally. Anatomical ROIs were defined based on labels in the Harvard-Oxford Cortical Structural Atlas in FSL for: right FG, left FG, right ATL, and left ATL. Functional ROIs were defined by centering a 20-mm-radius sphere on a MNI coordinate of interest. The spheres in bilateral FG were centered on the fusiform face area (FFA), which was defined individually for each participant with a functional localizer task (see Supplementary Methods). The spheres were centered on the FFA to ensure this region was included in the analysis, but extended well beyond the FFA. Cortex surrounding the

FFA was included because previous work suggests that it carries information about object category (Haxby et al., 2001), as well as viewpoint-independent representations of object exemplars, as revealed with fMRI adaptation (Pourtois, Schwartz, Seghier, Lazeyras, & Vuilleumier, 2005; Pourtois, Schwartz, Spiridon, Martuzzi, & Vuilleumier, 2009). The sphere in right ATL was centered on a coordinate where identity-specific patterns were previously found (Kriegeskorte et al., 2007), and the sphere in left ATL was centered on the mirror image of this coordinate. Thus, four functional ROIs were defined: right FG, left FG, right ATL, and left ATL. The number of voxels included in each of the eight ROIs can be found in Table S1 in Supplementary Material.

For inclusion in the pattern analyses, we selected voxels in each ROI that showed differential activity across identities in half of the data (the ‘training’ set). Specifically, we ran a one-way ANOVA over the 16 identities in the first run of each type (i.e. view-repeat or view-change) and selected voxels with $p < .05$ (one-tailed). This relatively liberal threshold was used to ensure a sufficient number of voxels for each ROI and participant. This feature selection was not biased because it was done blind to the other half of the data (the ‘test’ set). The correlation approach we employed is the simplest of pattern classifiers, in which the similarity of each test pattern was computed relative to the prototype training patterns from which feature selection had been applied. Moreover, we computed statistics based on the empirical null distribution from shuffling identity labels in the test set. Finally, we repeated the analyses without any feature selection (including all voxels in each ROI) and obtained similar results (see Supplementary Results). Because permutation tests are not conventional, we repeated all analyses using standard parametric significance tests (see Supplementary Methods and Results).

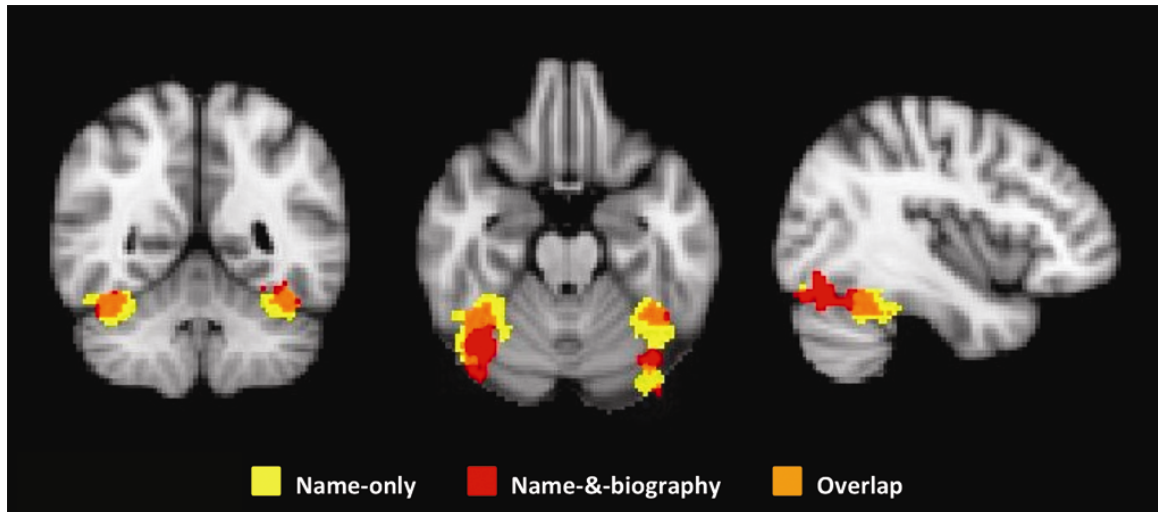


Figure 3.3. Identity adaptation. Clusters in a whole-brain analysis showing identity adaptation (identity-change > identity-same; corrected $p < .05$) for name-only (yellow) and name-and-biography identities (red). Voxels in orange exhibited adaptation in both social knowledge groups.

Results

Behavior

Accuracy and response time did not differ for name-and-biography versus name-only identities during the training or main tasks. Memory for which faces were associated with biographical information was near ceiling during the surprise post-test (see Supplementary Results for detailed behavioral results).

Identity Adaptation

fMRI adaptation was defined as greater signal change for identity-change versus identity-same trials, and was examined using whole-brain analyses. Three clusters showed identity adaptation ($p < .05$ corrected, MNI coordinates): right occipital fusiform gyrus (41, -65, -19; volume=14,032 mm³), left occipital fusiform gyrus (-38, -67, -19; volume=11,480 mm³), and the superior division of the left lateral occipital cortex (-29, -76, 30; volume=4,920 mm³). No clusters showed greater adaptation for name-and-

biography versus name-only identities. However, analyzing adaptation separately for these conditions revealed differences in the extent, location, and peak values (Figure 3.3).

Pattern Analyses

To create the identity maps, activation for the identity-same trials of a given identity was contrasted with a baseline of activation for identity-change trials. Removing this baseline eliminated correlations related to generic face processing, and as such, reduced correlation values drastically. For example, the mean correlation of maps belonging to different identities was higher before versus after removing the baseline in the FG ($r=.475$ vs. $r=.022$) and ATL ROIs ($r=.089$ vs. $r=.008$). In essence, the remaining correlations reflect the change in correlation from the generic pattern correlation of any two identities.

ROI	Identity	Interaction	Group
L FG anatomical	.004 *	.20	.43
R FG anatomical	<.001 *	.60	.30
L FG functional	.03 *	.34	.05 ‡
R FG functional	.03 *	.86	.004 *
L ATL anatomical	.85	.95 ‡	.32
R ATL anatomical	.04 *	.91	.15
L ATL functional	.82	.75	.13
R ATL functional	.71	.98 *	.03 *

Table 3.1. P-values for permutation tests of correlations. Identity: within-identity > within-group (one-tailed). Interaction: name-and-biography Identity vs. name-only Identity (two-tailed). Group: within-group > between-group (one-tailed). ‡ $p<.10$; * $p<.05$.

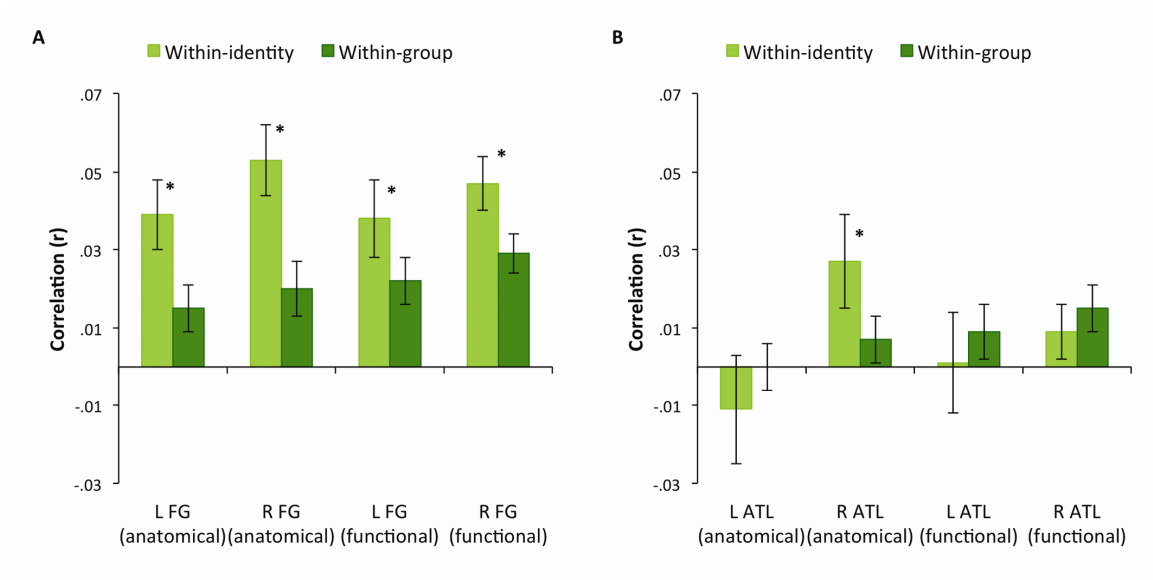


Figure 3.4. Pattern correlations for specific identities. Mean Pearson correlation coefficients for one identity with itself across runs (light green) and for two different identities that were either both name-and-biography or name-only identities (dark green) in FG (A) and ATL ROIs (B). Significance values were computed using permutation tests (Table 3.1). * $p < .05$. Error bars depict ± 1 s.e.m.

To test for identity-specific patterns, we compared within-identity and within-group pattern correlations (Figure 3.4; Table 3.1). Such representations were found in bilateral FG for both anatomical and functional ROIs, and in the right ATL anatomical ROI. The within-identity correlations were not significantly higher than the within-group correlations in the right ATL functional ROI, or in the left ATL anatomical or functional ROIs.

Learning biographical information about an individual might enhance familiarity and differentiate identity representations, leading to stronger identity correlations for the identities that were associated with vignettes. To test this, we compared the strength of within-identity correlations for the name-and-biography versus name-only faces using a two-tailed permutation test ('Interaction' in Table 3.1). Social knowledge did not increase identity correlations in any of the ROIs. In fact, biographical information decreased

identity correlations in the right ATL functional ROI and marginally in the left ATL anatomical ROI. However, neither of these ROIs represented specific identities (within-identity $\approx$ within-group), and thus it is difficult to interpret the finding that correlations were higher for name-only identities.

Although social knowledge did not increase the magnitude of identity correlations, it may have nevertheless altered the nature of the underlying representations. In support of this possibility, within-group correlations were greater than between-group correlations in the right FG functional ROI, right ATL functional ROI, and marginally greater in the left FG functional ROI (Figure 3.5). This was not true in any of the other ROIs. The results from the bilateral FG functional ROIs are particularly striking, since these same ROIs also carried identity representations (in comparison to the right ATL functional ROI, which did not).

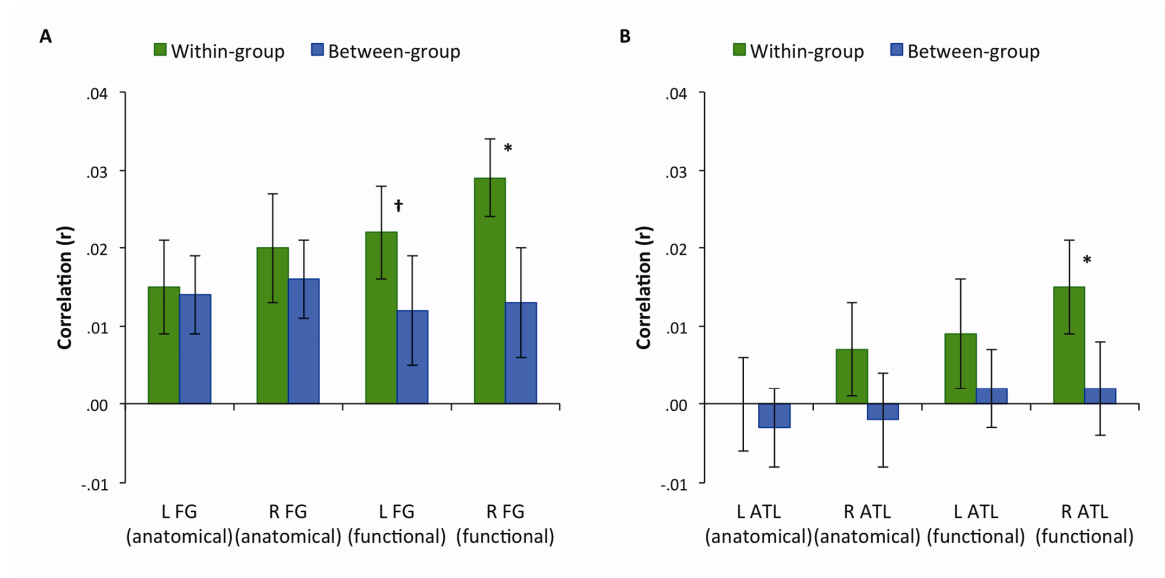


Figure 3.5. Pattern correlations for social knowledge groups. Mean Pearson correlation coefficients for two different identities that were either both name-and-biography or name-only identities (dark green) and for one name-and-biography identity with one name-only identity (blue) in FG (A) and ATL ROIs (B). Significance values were computed using permutation tests (Table 3.1). * $p < .05$. Error bars depict $\pm$ one s.e.m.

Discussion

We found evidence for individual-level information in the ventral temporal cortex, a region of the brain best known for its role in processing visual categories. In univariate analyses, large bilateral clusters in ventral temporal cortex showed fMRI adaptation when an identity was repeated versus changed. In multivariate analyses, activation maps in bilateral FG for a given identity were more highly correlated with themselves across runs than with maps for other identities. These two complementary techniques reveal the presence of facial identity information in the ventral temporal cortex at different spatial scales: fMRI adaptation indicates that identity-specific information exists within single voxels, while MVPA indicates that this information is also conveyed by distributed patterns of activity across multiple voxels.

We found evidence for individual-level information in the right ATL using pattern correlation analyses, but not using fMRI adaptation. The failure to find identity adaptation in the right ATL could indicate an absence of identity representations at the sub-voxel scale, or could reflect the response properties of the region. One previous study also failed to find a main effect identity adaptation in the ATL, but did report a correlation between the degree of adaptation and pre-experimental familiarity of participants with the famous faces in the study (Rotshtein et al., 2005). This suggests that, despite our training task, greater familiarity might be needed to observe identity adaptation in this region. Moreover, there is considerable BOLD signal dropout in this region (Devlin et al., 2000), and, perhaps because of this dropout, only recently was a face-selective patch in the human ATL reported (Rajimehr, Young, & Tootell, 2009). Nevertheless, such dropout did not prevent us from observing multivariate effects in the

ATL. Indeed, the finding of unique patterns for specific identities in the right but not left ATL replicates previous findings (Kriegeskorte et al., 2007; Nestor et al., 2011). This finding is also consistent with evidence that the right hemisphere is critical for face processing (Gazzaniga & Smylie, 1983; De Renzi, Perani, Carlesimo, Silveri, & Fazio, 1994; Le Grand, Mondloch, Maurer, & Brent, 2003), and may even be the source of face representations in the left hemisphere (Verosky & Turk-Browne, 2012).

We succeeded in finding robust identity-level patterns in ventral temporal cortex, despite the fact that this has proved difficult in the past (Kriegeskorte et al., 2007). One difference between our study and previous studies is that we created rich identity representations through training. There are two unique and potentially important aspects of our training: First, participants learned invariant identity representations by associating multiple viewpoints of each face with a name. Interestingly, the only other study to succeed in finding facial identity representations in posterior temporal cortex (Nestor et al., 2011) also employed multiple face exemplars for each identity — in their case, with different emotional expressions. Viewing faces over changes in visual properties may help create more invariant and robust representations (cf. Li & DiCarlo, 2008). The second unique aspect of our training was the use of many different identities. Previous studies have used two identities (Kriegeskorte et al., 2007) or four identities (Nestor et al., 2011), while we used 16 identities. Increasing the number of identities may have increased the difficulty of discriminating between them. Because our training task did not end until participants had learned to accurately discriminate identities, participants may have acquired more differentiated representations via perceptual learning (Goldstone, 1998). The fact that we found identity-level information in ventral temporal cortex

suggests that our training may have helped. Nevertheless, future work will be needed to directly compare identity representations for trained versus novel identities.

Beyond training, the manner in which we presented faces may have increased the sensitivity of our study. Previous studies presented one face per trial (Kriegeskorte et al., 2007; Nestor et al., 2011). In contrast, we presented pairs of faces in an fMRI adaptation design. Since pattern analyses were based on trials where identities repeated, it is possible that identity adaptation on these trials produced more distinct patterns. According to the sharpening model (Wiggs & Martin, 1998; Grill-Spector et al., 2006), adaptation occurs via local competition between selective and non-selective neurons, resulting in sparser and more selective neural activity for repeated objects. Trials containing both a trial-novel identity and its repetition may have evoked a more identity-specific response than a trial-novel identity would have on its own. Other models of adaptation, such as fatigue (e.g., McMahon & Olson, 2007), make the opposite prediction, in which case our adaptation design was detrimental to finding identity representations.

Social knowledge did not influence the magnitude of identity adaptation, nor did it increase the strength of pattern correlations for specific identities. However, it did alter the spatial distribution of adaptation and also grouped identity representations around a common pattern. Specifically, identity representations correlated more strongly with other identity representations that were associated with the same amount of biographical information. Consistent with the finding that even passive viewing of familiar faces engages regions throughout the brain more strongly than viewing of unfamiliar faces (Gobbini & Haxby, 2006, 2007; Todorov et al., 2007; Cloutier et al., 2011), we observed

this common pattern despite the fact that the name-matching task did not require explicit recall of biographical information.

While the presence of associated biographical information may account for the grouping effect, there are also other potential explanations. For instance, this effect could be due to changes in attention or to spontaneous categorization of faces into groups. The biographies included valenced information, and it is possible that simply viewing the faces associated with these biographies could have increased arousal. In addition, the fact that faces in each group were always paired with other faces from the same social knowledge condition may have served as a cue to group the faces together. Attention to faces (Wojciulik, Kanwisher, & Driver, 1998) and experimentally created group membership (Van Bavel, Packer, & Cunningham, 2011) have both been found to modulate activity in the FFA, indicating that this region is sensitive to top-down influence. While we did not observe a main effect of biographical information in our study, this could be due to either the name-matching task, which drew attention to the identity of faces in both groups, or to the rapid presentation of the images. Regardless, our data suggest that the FG is sensitive to the social context associated with faces, even when that context is not relevant to the task at hand. Future work could investigate how the content of the biographical information, such as learning that faces belong to a particular social group (e.g., political party, profession, etc.), alters face representations.

It has often been assumed that representations in ventral temporal cortex are unimodally visual, but a growing body of work suggests that this part of the brain may be sensitive to information from other sensory modalities (Von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005; Mahon, Anzellotti, Schwarzbach, Zampini, & Caramazza, 2009;

Van den Hurk, Gentile, & Jansma, 2011). For example, animal sounds presented auditorily elicit category-specific responses in ventral temporal cortex of congenitally blind adults, suggesting that visual experience is not necessary for this categorical organization (Mahon et al., 2009). In addition, the FFA responds more strongly to familiar versus unfamiliar voices during a speaker recognition task, further demonstrating that ventral temporal cortex can be co-opted for processing sound (Von Kriegstein et al., 2005). Finally, the FFA also responds differently to words describing categories of objects (e.g., pets, sports, etc.) and people, indicating a potential sensitivity to semantic content (Van den Hurk et al., 2011). Here, identities were randomly assigned to be paired with biographical information or not, such that faces with biographical information were not intrinsically more (visually) similar to each other. The finding that these faces were nevertheless represented more similarly in ventral temporal cortex suggests that this region is sensitive to information beyond mere appearance, including perhaps social information.

In conclusion, by creating rich identity representations through training, we succeeded in finding robust evidence for individual-level information in ventral temporal cortex in both univariate and multivariate analyses. These representations were sensitive to more than just visual properties, being influenced by the learning of biographical details about individuals. Together, our findings suggest that ventral temporal cortex is involved in individuating faces and open up exciting avenues for research into the role of learning in shaping face representations.

Conclusions and Future Directions

The first two parts of this dissertation showed how social knowledge influences our evaluation of novel faces, and the last part examined how identity is represented in the brain and how social knowledge shapes these representations. First, I demonstrated that rapidly learned affective associations can generalize to novel faces that resemble the familiar faces. Second, I showed that learning generalization based on physical similarity is a relatively automatic and powerful process, which likely influences face evaluation across a range of circumstances. Finally, I found that individual identities are tied to unique patterns of activity in ventral temporal cortex and that there are common patterns for faces associated with biographical information versus names alone. Taken together, these studies suggest that what we know about other people may actually change the way we perceive the faces of those around us. However, while these studies do show that social knowledge can influence face perception, they merely represent a first step in the broader project of examining how non-visual and visual information interact in our representations of other people.

For the work on learning generalization, I believe that an important direction for future research is to examine the types of learning that generalize. While the experiments presented in this dissertation show that learning need not be limited to significant others in order to generalize, they do not tell us about how the content of the learned information contributes to the observed learning generalization effect. In particular, the learned information in these studies is always valenced information regarding a person's actions, and it is not clear whether there is something special about valenced information or social information per se. For instance, would learning information about a person's

competence influence judgments about perceptually similar novel faces? Would learning non-social information about someone have the same effect as learning social information? While the current work demonstrates that rapidly learned affective associations contribute to our impressions of perceptually similar novel others, addressing these questions would allow us to speak to the specificity of the effect.

The findings in the pattern analysis experiment are more preliminary in the sense that the directions for future research follow more directly from them. First, although the fact that we were able to find unique patterns of activity for different identities in ventral temporal cortex suggests that the training task led to more robust identity representations, future work needs to verify the role of learning in shaping representations. For example, can we find changes in the strength of patterns associated with different identities as a result of experience? Second, while we were able to find common patterns for faces associated with biographical information versus names alone, it is not clear exactly what types of learning contribute to these patterns. For example, does the content of the learned information matter? While this experiment raises many questions, it is worth noting that it also provides us with a method for addressing these questions in the future.

The learning generalization studies in the first two parts of the dissertation and the imaging study in the last part of the dissertation both address the question of how learning influences perception of faces, but they do so in different ways. While the learning generalization studies help to tie changes in judgments of appearance to changes in perception, the imaging study examines how learning change the patterns of activity associated with different faces in visual areas of the brain. Although I have treated these two lines of work as separate, it is easy to imagine how they could be brought together in

the future. For instance, the methods established in the imaging study could be used to examine the neural basis of the learning generalization effect. As a first step, I could examine whether the patterns associated with different faces were sensitive to the physical similarity between faces. If this was found to be the case, I could examine the influence of learning behavioral information on the learned faces as well as on perceptually similar faces. It would also be possible to pit physical similarity and social similarity against each other: would we see different effects of physical and social similarity in different parts of the brain?

Thinking back to the introduction, different groups of researchers have focused on different aspects of person knowledge: vision scientists have focused on the processing of faces and bodies, while social psychologists have focused on the cognitive construal of others. However, visual and non-visual information are closely related: not only do we gain much of our information through vision, but the visual system itself was likely shaped by social constraints. In this dissertation, I explore one type of interaction between visual perception and social knowledge: I examine how social knowledge influences face perception. In doing so, I hope to have illustrated how both visual and non-visual knowledge contribute to our representations of others and how considering both sources of information opens up exciting avenues for future research.

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.
- Andersen, S. M., & Glassman, N. S. (1999). Activating transference without consciousness: Using significant-other representations to go beyond what is subliminally given. *Journal of Personality and Social Psychology*, 77, 1146-1162.
- 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.
- 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* (2nd ed., pp. 1–40). 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, & Affective Neuroscience*, 6, 572-581.
- 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.

- 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.
- 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: Academic Press.
- Cloutier, J., Kelley, W. M., & Heatherton, T. F. (2011). The influence of perceptual and knowledge-based familiarity on the neural substrates of face perception. *Social Neuroscience*, 6, 63-75.
- 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.
- De Renzi, E., Perani, D., Carlesimo, G. A., Silveri, M. C., & Fazio, F. (1994). Prosopagnosia can be associated with damage confined to the right hemisphere – an MRI and PET study and a review of the literature. *Neuropsychologia*, 32, 893-902.
- Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H. E., Matthews, P. M., & Tyler, L. K. (2000). Susceptibility-induced loss of signal: Comparing PET and fMRI on a semantic task. *NeuroImage*, 11, 589-600.

- 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.
- Gazzaniga, M. S., & Smylie C. S. (1983). Facial recognition and brain asymmetries: Clues to underlying mechanisms. *Annals of Neurology*, 13, 536-540.
- Gobbini, M. I., & Haxby, J. V. (2006). Neural response to the visual familiarity of faces. *Brain Research Bulletin*, 71, 76-82.
- Gobbini, M. I., & Haxby, J.V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, 45, 32-41.
- Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, 49, 585-612.
- Goldstone, R.L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology, General*, 130, 116-139.

- Goren, A., & Todorov, A. (2009). Two faces are better than one: Eliminating false trait associations with faces. *Social Cognition*, 27, 222-248.
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, 10, 14-23.
- Hamilton, D. L., Katz, L. B., & Leirer, V. O. (1980). Cognitive representation of personality impressions: Organization processes in first impression formation. *Journal of Personality and Social Psychology*, 39, 1050-1063.
- Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, 78, 837-852.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 28, 2425-2430.
- 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.
- Hills, P. J., & Lewis, M. B. (2007). Temporal limitation of Navon effect on face recognition. *Perceptual and Motor Skills*, 104, 501-509.
- 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.

- Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. *Journal of Experimental Social Psychology, 43*, 334-340.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513-541.
- Jacques, C. & Rossion, B. (2006). The speed of individual face categorization. *Psychological Science, 17*, 485-492.
- Johnson, M. K., Kim, J. K., & Risse, G (1985). Do alcoholic Korsakoff's syndrome patients acquire affective reactions? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 22-36.
- Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory, 17*, 577-596.
- 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.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*, 4302-4311.
- 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.

- Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 20600-20605.
- von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience*, 17, 367-376.
- Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2003). Expert face processing requires visual input to the right hemisphere during infancy. *Nature Neuroscience*, 6, 1108-1112.
- Levin, D. L., & Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces and inverted faces. *Perception and 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.
- Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representations in visual cortex. *Science*, 321, 1502-1507.
- Lundqvist, D., Flykt, A., & Ohman, A. (1998). *The Karolinska Directed Emotional Faces (KDEF)*, Stockholm: Department of Neurosciences Karolinska Hospital.
- Macrae, C. N., & Lewis, H. L. (2002). Do I know you? Processing orientation and face recognition. *Psychological Science*, 13, 194-196.

- Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron*, 63, 397-405.
- McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, 9, 605-610.
- McMahon, D. B., & Olson, C. R. (2007). Repetition suppression in monkey inferotemporal cortex: Relation to behavioral priming. *Journal of Neurophysiology*, 97, 3532-3543.
- Navon, D. (1977). Forest before the trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9, 353-383.
- Nestor, A., Plaut, D. C., & Behrmann, M. (2011). Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 9998-10003.
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the USA*, 105, 11087-11092.
- 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.
- 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.
- Pourtois, G., Schwartz, S., Seghier, M. L., Lazeyras, F., & Vuilleumier, P. (2005). Portraits or people? Distinct representations of face identity in the human visual cortex. *Journal of Cognitive Neuroscience*, 17, 1043-1057.
- Pourtois, G., Schwartz, S., Spiridon, M., Martuzzi, R., & Vuilleumier, P. (2009). Object representations for multiple visual categories overlap in lateral occipital and medial fusiform cortex. *Cerebral Cortex*, 19, 1806-1819.
- Quine, W. V. (1969). Natural kinds. In W. V. Quine. *Ontological Relativity and Other Essays* (pp. 114-138). New York: Columbia University Press.
- Rajimehr, R., Young, J. C., & Tootell, R. B. H. (2009). *Proceedings of the National Academy of Sciences of the United States of America*, 106, 1995-2000.
- 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.
- Rotshtein, P., Henson, R. N. A., Treves, A., Driver, J., & Dolan, R. J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience*, 8, 107-113.

- 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.
- Scott, L. S., & Monesson, A. (2009). The origin of biases in face perception. *Psychological Science*, 20, 676-680.
- Steyvers, M. (1999). Morphing techniques for generating and manipulating face images. *Behavior Research Methods, Instruments, & Computers*, 31, 359-369.
- Todorov, A., Gobbini, M. I., Evans, K. K., & Haxby, J. V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia*, 45, 163-173.
- 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 inference 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.
- Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 19514-19519.
- Turk-Browne, N. B., Scholl, B. J., & Chun, M. M. (2008). Babies and brains: Habituation in infant cognition and functional neuroimaging. *Frontiers in Human Neuroscience*, 2, 16.
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 43, 161-204.
- Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2011). Modulation of the fusiform face area following minimal exposure to motivationally relevant faces: Evidence of in-group enhancement (not out-group disregard). *Journal of Cognitive Neuroscience*, 23, 3343-3354.
- Van den Hurk, J., Gentile, F., & Jansma, B. M. (2011). What's behind a face: Person context coding in fusiform face area as revealed by multivoxel pattern analysis. *Cerebral Cortex*, 21, 2893-2899.

- Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science*, 21, 779-785.
- Verosky, S. C., & Todorov, A. (submitted). When physical similarity matters: Mechanisms underlying affective learning generalization to the evaluation of novel faces.
- Verosky, S. C., Todorov, A., & Turk-Browne, N. B. (submitted). Representations of individuals in ventral temporal cortex defined by faces and biographies.
- Verosky, S. C., & Turk-Browne, N. B. (2012). Representations of facial identity in the left hemisphere require right hemisphere processing. *Journal of Cognitive Neuroscience*, 24, 1006-1017.
- Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, 8, 227-233.
- Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Modulation of activity in the fusiform face area by covert attention: An fMRI study. *Journal of Neurophysiology*, 79, 1574-1579.
- 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. & 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.

Footnotes

1. Including these three participants did not affect the results.
2. Although previous studies have found an effect of reporting global versus local shape of a Navon letter on face recognition (Macrae & Lewis, 2002; Hills & Lewis, 2007), we did not replicate this effect in our study. While there are many potential explanations for this null effect, one stands out: participants in the current study had more experience with the faces they were asked to recognize than did participants in previous studies. Research on the Navon effect grew out of research examining eyewitness memory and, as a result, initial exposure to the learned faces in the previous studies has tended to be brief, often on the order of seconds. In contrast, participants in the current study saw each face several times and were tested on their ability to differentiate between faces. People are better at recognizing familiar versus unfamiliar faces across a variety of changes in visual properties (Johnston & Edmonds, 2009) and therefore it seems possible that this additional experience helped create face representations that were less susceptible to changes in processing strategy.

Supplementary Material for Part 1

Methods

Preliminary study 1. We morphed 18 pairs of faces in increments of 10 percent. In addition, morphs containing 35 and 65 percent of the base faces were included because the greatest change in perceived identity was expected to occur close to these positions. On each trial, the morphs were presented in the middle of the screen following a 500-millisecond fixation cross, with the endpoint faces on either side. Participants were asked to categorize the morph as looking like one or the other endpoint face. The morph was presented until the participant's response. The 13 morphs for each pair of faces were blocked together. There were 18 blocks of faces for a total of 234 trials. The order of the morphs in each pair and the order of the pairs were both randomized.

Preliminary study 2. As in the evaluation phase of the main experiment, half of the 36 morphed faces were shown at one level of morphing and half were shown at the other and this was counterbalanced across participants. Each trial began with a 1000-millisecond fixation cross, followed by a face. Each face remained on the screen until participants categorized it as learned, similar to learned, or new. Participants responded using the number keys 1 (learned), 2 (similar to learned), and 3 (novel). When participants responded "learned", the face remained on the screen and they were asked to further indicate whether it was previously paired with negative, neutral, or positive information. Similarly, when participants responded "similar to learned", they were asked to indicate whether the face that the current face resembled was previously paired with

negative, neutral or positive information. Participants responded to the second question using the number keys 4 (negative), 5 (neutral), and 6 (positive).

Additional Results for Preliminary Study 2.

Categorization of learned faces. In the learning phase of the experiment, 33% of participants were able to correctly identify which faces had been paired with which type of behavior after two test rounds, more than 90% were able to do so after four test rounds, and 100% were able to do so after five test rounds.

In the categorization phase, participants categorized the learned faces as learned most of the time (Figure 1.2B in main text; $M=.91$, $SD=.14$). After correctly categorizing the learned faces as learned, participants were quite good at identifying the valence of the behavior the face was previously paired with (Figure S1; $M=.91$, $SD=.12$). Thus, participants were able to recognize both that the face was learned and the valence of the behavior associated with that face most of the time ($M=.83$, $SD=.18$). When participants did not categorize the learned faces as learned, they categorized them as similar to learned ($M=.09$, $SD=.14$). Even on these trials, participants were still able to identify the valence of the behavior associated with the original face most of the time ($M=.83$, $SD=.26$).

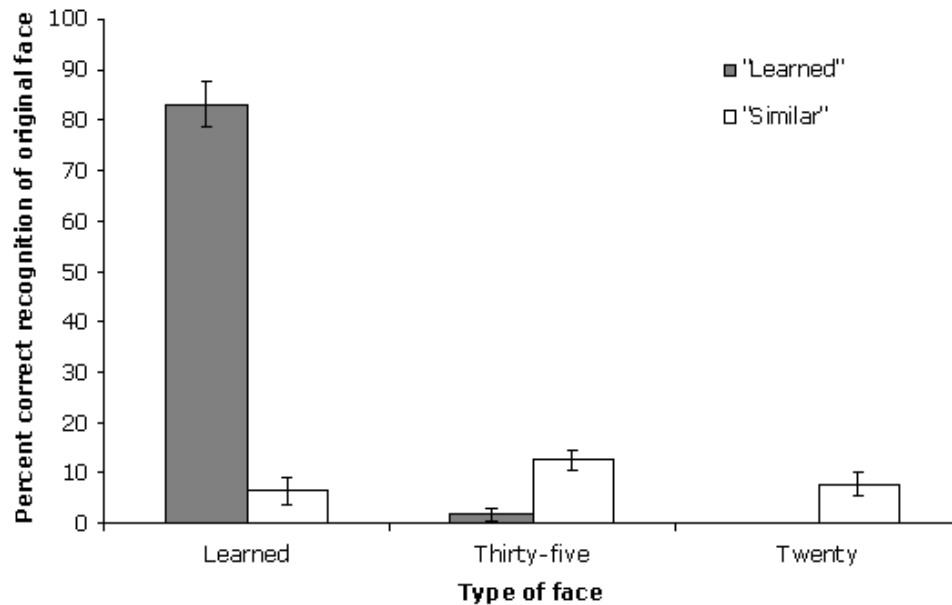


Figure S1. Correct identification of similarity to the learned faces and valence of behavior by type of face ($N=15$). Error bars represent standard error of the mean.

Categorization of morphs and novel faces. Participants were slightly more likely to categorize the novel faces ($M=.15$, $SD=.11$) and the 35% morphs ($M=.16$, $SD=.09$) as learned or as similar to learned than the 20% morphs (Figure 1.2B in main text; $M=.10$, $SD=.09$), $F(2, 28)=2.97$, $p=.07$, $\eta_p^2=.18$. However, the likelihood of categorizing faces as learned versus similar to learned did not depend on the type of face, $F<1$ for the interaction of type of face and categorization. The increased likelihood of categorizing novel faces versus 20% morphs as learned or similar to learned is probably due to the fact that the faces used to create the morphs were selected based on their low similarity to the learned faces. Because the 20% morphs contain more of these original faces than the 35% morphs, it is not surprising that participants were less likely to categorize the 20% morphs as learned or similar to learned than the 35% morphs.

Importantly, the categorization of the 35% morphs was indistinguishable from the categorization of novel faces for all three levels of categorization, $t_s < 1$.

Participants only rarely mistook the 20% morphs ($M=.004$, $SD=.01$) or the 35% morphs ($M=.04$, $SD=.06$) for learned faces. For the 20% morphs, only one participant made this mistake and this person subsequently failed to identify the valence of the behavior associated with the learned face. For the 35% morphs, participants were able to identify the types of behaviors associated with the learned faces less than half of the time ($M=.43$, $SD=.53$), meaning that they could only have mistaken the 35% morphs for particular learned faces on less than 2% of trials.

On trials where participants correctly categorized the 20% morphs ($M=.20$, $SD=.17$) and the 35% morphs ($M=.28$, $SD=.17$) as similar to the learned faces, they were still not particularly good at identifying the valence of the behavior associated with the original face (for 20% morphs: $M=.38$, $SD=.16$; for 35% morphs: $M=.51$, $SD=.27$). That is, participants correctly identified both the similarity to the learned faces and the valence of the behavior associated with these faces on less than 14% of the trials for the 35% morphs and on less than 8% of the trials for the 20% morphs.

Supplementary Material for Part 3

Supplementary Methods

Main task. *Pattern analysis with standard significance tests.* Although non-parametric permutation tests are beneficial in not making assumptions about the null distribution of pattern correlations, they are less conventional and more conservative than parametric statistical tests. To further support our conclusions, we also tested for identity- and group-level information using standard statistical tests. Prior to these analyses, correlations were transformed to normally-distributed Fisher's z scores. To test for unique patterns associated with specific identities, the within-identity and within-group correlations were entered into a 2 (within-identity, within-group) $\times$ 2 (name-and-biography, name-only) repeated-measures ANOVA. To test for distinct patterns associated with each social knowledge group, within-group correlations were compared to between-group correlations using a paired samples t-test.

Localizer task. After the main task, participants completed two runs of a functional localizer task to define the FFA. Each run contained color photographs of faces (from a different set than those used in the main task) and scenes in separate blocks. Participants indicated whether faces were male or female and whether scenes were indoor or outdoor. Each block contained 12 images, presented for 500 ms each followed by a 1000-ms ISI, for a duration of 18 s (followed by 12 s of fixation). Images were presented in the center of the screen, and subtended $5 \times 5^\circ$. Each run contained six face blocks and six scene blocks, shown in alternating order (starting category was counterbalanced across participants), and lasted 369 s (including 9 s of discarded acquisitions).

Localizer data were preprocessed using the same parameters as the main task. For each run, boxcar functions lasting 18 s were placed at that onset of each face or scene block and convolved with a double-gamma hemodynamic response function, and these two regressors were entered in a GLM. The resulting parameter estimates were combined across the two localizer runs within participant using a fixed-effects GLM. To determine FFA coordinates, parameter estimates for face and scene blocks were contrasted, and the peak voxel in an anatomically-restricted region of the fusiform gyrus was selected (e.g., Verosky & Turk-Browne, 2012).

Memory test. After completing the scan session, participants were given a surprise test of memory for the biographical information presented with the faces. For each identity, participants were shown the image of the person facing directly forward and were asked to indicate whether they had learned any biographical information about the person using the number keys 1 ("yes") and 2 ("no"). If they indicated that they had learned biographical information about the person, they were asked to further indicate whether it was primarily negative or positive in tone using the number keys 3 ("negative") and 4 ("positive"). Each trial began with 1 s of fixation and each image remained onscreen until response. The 16 identities were shown in a random order.

Supplementary Results

Behavior. Training. During the first test block of the training task (i.e., after one round of training), accuracy in identifying the face that went with each name was significantly above chance ($M=69.58\%$, $SD=13.91\%$, chance= 33.33% ; $t(19)=11.76$, $p<.001$). During the second test block (i.e., after another round of training), accuracy

improved ($M=90.73\%$; $SD=10.22\%$; vs. chance, $t(19)=25.27$, $p<.001$). Nine out of the twenty participants reached the criterion of 95% correct during this block and did not complete any further rounds of the task. During the third test block, accuracy improved further ($M=96.78\%$, $SD=3.00\%$; vs. chance, $t(10)=70.53$, $p<.001$), and nine more participants reached criterion. During the fourth test block, the remaining two participants reached criterion ($M=98.96\%$, $SD=1.04\%$). Across all test blocks, participants were no more accurate in identifying the face that went with each name for the name-and-biography ($M=85.07\%$, $SD=7.83\%$) versus name-only identities ($M=85.00\%$, $SD=7.89$; $t<1$), and no faster to respond for the name-and-biography ($M=1322$ ms; $SD=431$ ms) versus name-only identities ($M=1266$ ms; $SD=363$ ms; $t(19)=1.19$, $p=.25$). Accuracy and response time for name-and-biography versus name-only identities did not differ for any of the individual test blocks either (accuracy: $ps>.13$; response time: $ps>.12$).

Main task. Accuracy on the main task was near ceiling ($M=97.80\%$, $SD=2.11\%$). There was no main effect of whether a face was associated with biographical information ($F<1$), but there was a marginally significant main effect of identity, such that participants were slightly more accurate for the identity-same ($M=97.99\%$; $SD=1.98\%$) than for the identity-change trials ($M=97.62\%$, $SD=2.31\%$; $F(1,19)=3.61$, $p=.07$). This main effect was qualified by a marginally significant interaction with whether a face was associated with biographical information ($F(1,19)=3.43$, $p=.08$). For the name-only identities, participants were significantly more accurate for the identity-same ($M=98.24\%$, $SD=1.98\%$) than the identity-change trials ($M=97.42\%$, $SD=2.79\%$; $t(19)=2.45$, $p=.03$). However, for the name-and-biography identities, there was no

difference between identity-same ($M=97.73\%$, $SD=2.35\%$) and identity-change trials ($M=97.81\%$, $SD=2.26\%$; $t<1$).

In response times (RTs), there were no main effects of biographical information ($F(1,19)=1.87$, $p=.19$) or identity ($F<1$). However, there was a marginally significant interaction between biographical information and identity ($F(1,19)=3.45$, $p=.08$). For the name-only identities, participants were numerically faster for the identity-same ($M=713$ ms, $SD=92$ ms) than for the identity-change trials ($M=725$ ms, $SD=105$ ms; $t(19)=1.48$, $p=.16$). For the name-and-biography identities, participants were numerically slower for the identity-same ($M=731$ ms, $SD=99$ ms) than for the identity-change trials ($M=730$ ms, $SD=98$ ms; $t<1$). This analysis was based on correct trials only, but including RTs from incorrect trials did not change the pattern of results.

Memory test. Participants accurately identified which faces were associated with biographical information ($M=97.81\%$, $SD=5.08\%$; vs. chance=50%: $t(19)=86.12$, $p<.001$), and whether that information was negative or positive ($M=94.37\%$, $SD=8.58\%$; vs. chance=50%: $t(19)=49.20$, $p<.001$), despite never having been tested on this information during training.

Region	Number of voxels	Within-identity	Within-group	Between-group
L FG anatomical	2894	.029 (.008)	.013 (.006)	.010 (.006)
R FG anatomical	3151	.029 (.008)	.010 (.005)	.009 (.006)
L FG functional	4169	.026 (.009)	.015 (.006)	.002 (.006)
R FG functional	4169	.028 (.008)	.021 (.004)	.010 (.005)
L ATL anatomical	4032	-.002 (.010)	.003 (.006)	.002 (.004)
R ATL anatomical	4012	.011 (.009)	.005 (.005)	-.001 (.004)
L ATL functional	4169	.001 (.008)	.004 (.004)	.004 (.005)
R ATL functional	4169	.004 (.007)	.010 (.004)	.000 (.004)

Table S1. Mean (and standard error) Pearson correlation coefficients based on all voxels within each ROI (without feature selection). Within-identity: correlation of each identity with itself across runs. Within-group: correlation of two different identities that were either both name-and-biography or name-only. Between-group: correlation of one name-and-biography and one name-only identity.

Pattern analyses. Robustness of findings. To summarize, the observed effects were slightly stronger when correlations were computed over voxels that were sensitive to identity in half of the data (Figures 3.4 and 3.5 in the main text) versus when computed over all voxels in each ROI (Table S1), and the permutation tests were slightly more sensitive than standard significance tests. Despite these small differences, the overall pattern of results was consistent across the four analyses: permutation tests with (Table 3.1) and without (Table S2) feature selection; standard tests with (Table S3) and without (Table S4) feature selection.

We found evidence for identity-specific patterns in the bilateral FG anatomical ROIs across all four analyses. In the bilateral FG functional ROIs and the right ATL

anatomical ROI, this identity effect (permutation: within-identity vs. shuffled within-group; standard: main effect of within-identity vs. within-group) was somewhat more sensitive to the analysis method. For the right FG functional ROI, the identity effect was significant with but not without feature selection in both permutation and standard tests. Similarly, for the left FG functional ROI, the identity effect was significant in the permutation test with feature selection but only marginally significant without, and was marginally significant in the standard test with feature selection but non-significant without. For the right ATL anatomical ROI, the identity effect was significant in the permutation test with feature selection but not without, and was marginally significant in the standard test with feature selection but again not without.

ROI	Identity	Interaction	Group
L FG anatomical	.01 *	.35	.27
R FG anatomical	.001 *	.78	.48
L FG functional	.07 ‡	.21	.005 *
R FG functional	.20	.95	.007 *
L ATL anatomical	.74	.94	.46
R ATL anatomical	.19	.97 ‡	.10
L ATL functional	.67	.75	.50
R ATL functional	.80	.94	.01 *

Table S2. P-values for permutation tests of correlations without feature selection. Identity: within-identity > within-group (one-tailed). Interaction: name-and-biography Identity vs. name-only Identity (two-tailed). Group: within-group > between-group (one-tailed). ‡ $p < .10$; * $p < .05$.

ROI	Identity		Interaction		Group	
	<i>F</i> (1,19)	<i>p</i>	<i>F</i> (1,19)	<i>p</i>	<i>t</i> (19)	<i>p</i>
L FG anatomical	7.89	.006 *	.59	.45	.15	.44
R FG anatomical	9.55	.003 *	.08	.78	.60	.28
L FG functional	2.09	.08 ‡	.18	.67	1.33	.10 ‡
R FG functional	5.20	.02 *	.86	.36	2.81	.006 *
L ATL anatomical	.79	.19	1.94	.18	.52	.30
R ATL anatomical	2.92	.05 ‡	1.91	.18	1.13	.14
L ATL functional	.57	.23	.54	.47	.83	.21
R ATL functional	.61	.22	6.86	.02 *	1.71	.05 ‡

Table S3. Standard significance tests of correlations with feature selection. Identity: within-identity > within-group (one-tailed). Interaction: name-and-biography Identity vs. name-only Identity (two-tailed). Group: within-group > between-group (one-tailed). The main effect of Biographical Information (name-and-biography vs. name-only) is not shown and was not significant in any region (two-tailed; *ps* >.31). ‡ *p* < .10; * *p* < .05.

The test of whether the identity effect was influenced by biographical information was significant in the right ATL functional ROI in all cases except the permutation test without feature selection. As previously described, this interaction effect (permutation: name-and-biography within-identity vs. name-only within-identity; standard: interaction of within-identity/within-group vs. name-and-biography/name-only) was in an unexpected direction, with a stronger identity effect for the name-only identities than the name-and-biography identities. However, given the overall absence of an identity effect in the right ATL functional ROI, the interaction effect should be interpreted with caution. Regardless, across all ROIs, biographical information did not *increase* the strength of identity correlations.

ROI	Identity		Interaction		Group	
	<i>F</i> (1,19)	<i>p</i>	<i>F</i> (1,19)	<i>p</i>	<i>t</i> (19)	<i>p</i>
L FG anatomical	5.82	.01 *	.27	.61	.52	.30
R FG anatomical	4.27	.03 *	.84	.37	.05	.48
L FG functional	1.50	.24	.62	.22	1.92	.03 *
R FG functional	.59	.23	3.49	.08 ‡	2.25	.02 *
L ATL anatomical	.30	.30	2.17	.16	.09	.46
R ATL anatomical	.76	.20	2.70	.12	1.06	.15
L ATL functional	.13	.36	.57	.46	.00	.50
R ATL functional	1.05	.16	6.38	.02 *	1.69	.05 ‡

Table S4. Standard significance tests of correlations without feature selection. Identity: within-identity > within-group (one-tailed). Interaction: name-and-biography Identity vs. name-only Identity (two-tailed). Group: within-group > between-group (one-tailed). The main effect of Biographical Information (name-and-biography vs. name-only) is not shown and was not significant in any region (two-tailed; *ps* >.59). ‡ *p* < .10; * *p* < .05.

Finally, we found consistent evidence for group-level information in the bilateral FG functional ROIs across analyses. For the right FG functional ROI, this group effect (permutation: within-group vs. shuffled between-group; standard: planned comparison of within-group vs. between-group) was significant across all four analyses. For the left FG functional ROI, the group effect was marginally significant in the permutation test with feature selection and significant without, and was marginally significant in the standard test with feature selection and again significant without. Finally, for the right ATL functional ROI, the group effect was significant with and without feature selection in the permutation test, and was marginally significant with and without feature selection in the standard test.

Appendix A: Biographical vignettes used in Part 2, Experiment 1

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 coworkers, 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.

Appendix B: Example biographical vignettes for Part 3

Negative example: Jessica was laid off from her job in a department store because she was caught stealing. She would sneak expensive items of clothing out of the store in her purse and then she would return them at a different branch of the store for cash. She also bought several pieces of jewelry at a discount price on employee discount day and then returned them for full price. While she worked at the store, people's belongings would often go missing from the employee break room; now that she is gone, this no longer happens. On the one hand, people were surprised to learn that she had been doing this: her job in the store paid well and the amount of money she made through the returns was small in comparison. On the other hand though, she had never been popular among her coworkers. After losing her job, she decided to comfort herself by taking a short trip to Florida. She has been spending her days on the beach – trying to think of clever ways to make money – and her nights in the club.

Positive example: Tiffany is a good role model for her students. She began her job as a high school teacher three years ago and since then test scores have gone up and more students have gone to college. She teaches chemistry – hardly the most glamorous subject – but somehow she manages to get all of the students involved. Not only is she good at explaining difficult concepts, but she is also good at conveying her enthusiasm for the subject. She genuinely enjoys spending time with her students and sometimes she will organize weekend trips to take them to places – like the Natural History Museum -

where they otherwise would not think to go. In her spare time, she also enjoys being active and doing things like hiking and going to the gym. She has never thought of herself as much of a runner, but recently she and her husband started training for a half marathon. As the marathon gets closer, they will go on longer and longer runs. Although the first few days have been tough going, she keeps imagining how amazing it will feel to have done this.