Social Cognition, Vol. 33, No. 5, 2015, pp. 368–386

THE ROBUSTNESS OF LEARNING ABOUT THE TRUSTWORTHINESS OF OTHER PEOPLE

Virginia Falvello and Michael Vinson
Princeton University

Chiara Ferrari
Princeton University; University of Pavia, Italy

Alexander Todorov
Princeton University

A single behavioral statement about a person is sufficient to form trustworthiness associations with the person's face. In three experiments, we tested the limits of these associative processes. Participants were presented with 100 to 500 face-behavior pairs (intermixed with place images so that all participants saw 500 images and descriptions) and then judged a subset of the faces. Participants rated faces paired with positive behaviors as more trustworthy than faces paired with negative behaviors. This effect was as strong after 100 as after 400 face-behavior pairs (Experiment 1). In subsequent experiments, we tested whether similar associative effects could be obtained for places (Experiment 2) and competence-related behaviors (Experiment 3). Although we expected that the associative effects would be stronger for faces than places and for morality- than competence-related behaviors, we obtained similar effects. The findings suggest that trustworthiness associations are based on global affect-based inferences.

Much of the recent research on first impressions focuses on trait judgments from facial appearance (for a review see Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). This research shows that people can rapidly and effortlessly evaluate others based on their facial appearance alone, and that the perceptual basis of these evaluations could be understood. Different traits evaluations are highly corre-

Address correspondence to Alexander Todorov, Peretsman-Scully Hall, Psychology Department, Princeton University, Princeton, NJ 08540; E-mail: atodorov@princeton.edu

lated with each other and are reducible to two or three underlying dimensions (McAleer, Todorov, & Belin, 2014; Oosterhof & Todorov, 2008; Sutherland et al., 2013; Todorov, Said, Engell, & Oosterhof, 2008). The first dimension, which accounts for most of the variance of evaluations, is about the perceived goodness or badness of faces. Importantly, the judgment closest to this dimension is trustworthiness. Not only do people infer trustworthiness from faces after extremely brief exposures (Todorov, Loehr, & Oosterhof, 2010; Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006), but they also act on these inferences trusting more trustworthy looking partners in economic games (Chang, Doll, van't Wout, Frank, & Sanfey, 2010; Ewing, Caulfield, Read, & Rhodes, 2014; Rezlescu, Duchaine, Olivola, & Chater, 2012; Schlicht, Shinsuke, Camerer, Battaglia, & Nakayama, 2010; Stirrat & Perrett, 2010; Tingley, 2014; van't Wout & Sanfey, 2008). The importance of trustworthiness judgments converges with recent work showing that inferences of moral character dominate person evaluation (Goodwin, Piazza, & Rozin, 2013; Uhlmann, Pizarro, & Diermeier, 2015).

However, appearance is not the only basis of trustworthiness judgments. People highly value information about others and are remarkably good at learning such information. According to some estimates, approximately two-thirds of conversation time between acquaintances is spent "gossiping" or discussing social topics (Dunbar, Duncan, & Marriott, 1997). This gossiping has reputational consequences. Research on spontaneous trait inferences shows that people automatically make person inferences after reading about a single behavioral act (Uleman, Newman, & Moskowitz, 1996) and that these inferences are associated with the person's face (Bliss-Moreau, Barrett, & Wright, 2008; Todorov & Uleman, 2002, 2003, 2004). Given people's motivations to figure out the moral character of others (Uhlmann et al., 2015), inferences of trustworthiness from others' behaviors could be particularly strongly associated with their faces. Previous work has already shown that people make trustworthiness judgments from a single good or bad behavior (Todorov & Olson, 2008). This learning even generalizes to the evaluation of novel faces: novel faces resembling faces associated with untrustworthy behaviors are judged as less trustworthy than novel faces resembling faces associated with trustworthy behaviors (Verosky & Todorov, 2010, 2013).

In this research, we tested the robustness of forming trustworthiness associations between faces and behaviors, presenting participants with as many as 500 faces and behaviors. Past studies have typically shown participants between 30 and 120 faces paired with behavioral information, and Todorov and Uleman (2002, 2004) found that participants were able to associate trait inferences from behaviors with faces even when viewing as many as 120 such pairings. In fact, across their experiments, there was little evidence that the strength of associations between faces and inferred traits decreased with the increase in the number of face-behavior pairs. However, no studies have examined the ability to form and retrieve trustworthiness associations for larger sets of faces and behaviors. In the current experiments, we varied the number of face-behavior pairs (ranging from 100 to 500) to test whether the strength of these associations decreases as the number of face-behavior pairs increases.

Studies on social network size suggest an approximate number of trustworthiness associations people can reliably form and retrieve. In a study designed to measure the size of acquaintanceship networks, Killworth, Bernard, and McCarty (1984) estimated that participants knew approximately 134 people of whom they felt comfortable asking favors. Dunbar's social brain hypothesis estimates the size of the optimal human social network to be approximately 150 individuals (Dunbar, 1992, 1993, 1998). Extrapolating from these findings, one hypothesis is that the ability to form and retrieve trustworthiness associations will decrease when viewing more than 200 faces with behavioral information. On the other hand, human memory for images is phenomenal (Standing, 1973; Standing, Conezio, & Haber, 1970), and the ability to retrieve affective associations about others appears not to depend on explicit memory (Johnson, Kim, & Risse, 1985; Todorov & Olson, 2008; Tranel & Damasio, 1993). Because it is possible to retrieve affective associations with faces without remembering specific information about the faces, the maximum number of such associations that can be formed may not be limited by social network size, which does involve explicit memory of names and relationships.

In three experiments, we manipulated the number of face-and-behavior pairs to which participants were exposed to test whether the increase in the number of these pairs decreases the strength of affective associations with faces. Specifically, after participants were presented with the face-behavior pairs, they judged a subset of the faces. Consistent with many prior studies (Bliss-Moreau et al., 2008; Carlston & Skowronski, 1994; Carlston, Skowronski, & Sparks, 1995; Cassidy & Gutchess, 2014; Crawford, Skowronski, Stiff, & Scherer, 2007; Todorov & Uleman, 2002, 2003, 2004; Verosky & Todorov, 2010, 2013), we expected that faces that were paired with positive or neutral behaviors would be judged as more trustworthy than faces paired with negative behaviors. This was the case across three experiments. In Experiment 1, the effect of behavior was as strong after 100 as after 400 face-behavior pairs. In Experiment 2, we replicated this finding and tested whether the effects for faces are stronger than similar associative effects for places. In Experiment 3, we tested whether the associative face effects are stronger for morality- than competence-related behaviors. We also explored the relation between the ability to recognize faces and the ability to judge them according to the valence implications of the associated behaviors. To the extent that face recognition is a necessary precondition for retrieval of trustworthiness associations about the face, we expected to find a positive correlation between recognition performance and the difference in judgments of faces associated with positive behaviors and faces associated with negative behaviors.

EXPERIMENT 1

The objective of this experiment was to test whether the strength of trustworthiness associations between faces and behaviors decreases as the number of faces and behaviors increases. Four groups of participants were shown 100, 200, 300, or 400 faces paired with behavioral descriptions that were positive, neutral, or negative in "goodness." The face images and behaviors were intermixed with place images and descriptions so that participants in all groups saw 500 images. Afterward, participants were shown 60 of the viewed faces without any behaviors, and rated the faces on trustworthiness.

METHODS

Participants. Eighty Princeton undergraduate students (34 male) participated in exchange for a payment of $12.

Materials. Black and white, forward-facing, un-posed facial photos of two hundred male and two hundred female individuals were selected from the FERET Grayscale Database CD 1 (Phillips, Moon, Rizvi, & Rauss, 1999). The selected photos varied somewhat in lighting conditions and though most faces exhibited neutral expressions, some were smiling. Four hundred photographs of outdoor scenes were selected from the collection of place images used by Konkle, Brady, Alvarez, and Oliva (2010). To obtain the desired number of place photos, all photos from seven scene categories were used: mountains, streets, underwater images, beaches, coasts, castles, and gardens. This database was originally created to explore the ability to detect minute differences in pictures and contained some photographs of the same places from different angles or zooms. Since the purpose of the place photos in this experiment was to balance the number of photographs viewed by different participants and the place photos were not rated, all place photos were included in this experiment.

Three hundred behavioral descriptions were taken from Fuhrman, Bodenhausen, and Lichtenstein (1989), which provides ratings of behavioral statements on traits such as kindness, intelligence, goodness, and normality. Out of the 400 statements evaluated by Fuhrman, Bodenhausen, and Lichtenstein, we selected the 100 most positive, 100 most negative, and 100 most neutral behavioral statements based on their "goodness" ratings, as these ratings seemed most likely to predict perceived trustworthiness. The 100 positive statements ranged in "goodness" from a mean of 6.95 to 9.08 on an 11-point scale. The 100 neutral statements ranged from 4.74 to 6.23, and the 100 negative statements ranged from 0.51 to 3.28. Because 400 unique behavioral statements were needed in total, we created an additional 33 positive, 33 negative, and 34 neutral behavioral statements. However, only behaviors with ratings from Fuhrman, Bodenhausen, and Lichtenstein (1989) were rated in the trustworthiness-rating task described below. Examples of typical negative, neutral, and positive statements include "he intentionally swerved his car to hit a squirrel," "he took a nap one afternoon," and "he helped push someone's car out of a snowbank," respectively. Each face photo was randomly assigned to a unique behavioral statement. All participants viewed equal numbers of male and female faces, and as equal a number of faces paired with positive, neutral, or negative behavioral descriptions as each condition allowed. Photos of places were accompanied by one of the following generic phrases: "good place," "neutral place," or "bad place."

Procedures. Each participant was randomly assigned to one of four conditions: 100 faces and 400 places, 200 faces and 300 places, 300 faces and 200 places, or 400 faces and 100 places. Thus, every participant viewed 500 photos total in a 4 (group-size condition) by 3 (valence) mixed factorial design.

Participants were told that they would be viewing photographs of faces and places that would be paired with negative, neutral, or positive descriptions. Participants were instructed that each face would only be paired with one type of behavior and that participants should form impressions of the people pictured by imagining the people actually performing the behaviors. The 500 photos were shown on the screen with their unique descriptions in consecutive sequence for four and a half seconds each. A 500 ms fixation cross was shown between photos. The order in which the photos were presented was randomized for every participant.

After viewing the 500 photographs and descriptions, participants were shown 60 face photographs from the impression formation task without their accompanying behaviors, in random order. Participants rated how trustworthy each face appeared to them on a scale from 1 (not at all trustworthy) to 9 (very trustworthy). Participants were instructed that there were no right or wrong answers and to use their intuitive, "gut" judgment. All participants rated the same 60 faces regardless of group-size condition. The 60 rated faces were never presented in the first or last 100 photos during the impression-formation task, but were randomly presented within the middle 300 photos to ensure that primacy and recency effects did not act as confounding variables on ratings of trustworthiness. Additionally, the rated faces were equally weighted in terms of gender, so that 30 male and 30 female faces were rated, and valence-behavior groups, so that 20 faces that had been paired with positive behaviors, 20 faces that had been paired with neutral behaviors, and 20 faces that had been paired with negative behaviors were rated.

After the trustworthiness ratings, participants were shown 20 more faces, 10 of which had been presented during the impression-formation task but not rated in the trustworthiness ratings task and 10 of which were novel and had not been presented previously at any point in the experiment. Participants were simply asked to indicate whether or not they remembered seeing each face during the impression-formation task in a forced-choice task. The faces were equally weighted in terms of gender and were shown in randomized order for each participant.

RESULTS AND DISCUSSION

Participants rated faces paired with positive and neutral behaviors as more trustworthy than those paired with negative behaviors (see Table 1), as confirmed by a significant main effect of valence $F(2, 152) = 61.62$, $p < .001$, $\eta^2 = .45$. There were no other significant effects (Fs < 1.02, ps > .42 for the main effect of group size, and the interaction of group size and valence). Post-hoc comparisons (Bonferroni correction applied) showed that faces paired with positive behaviors were rated as more trustworthy ($M = 5.42$, $SD = 0.58$) than those paired with neutral behaviors ($M = 5.25$, $SD = 0.55$), $t(79) = 2.97$, $p < .014$, $d = .33$. In turn, the latter were rated as more

TABLE 1. Means (SD) of Trustworthiness Judgments of Faces as a Function of the Valence of the Behaviors Associated with the Faces and the Number of Faces-and-Behaviors Pairs

	Valence of behavior		
Faces and behaviors	Positive	Neutral	Negative
100	5.28 (0.53)	5.10 (0.48)	4.53 (1.12)
200	5.41 (0.73)	5.28 (0.66)	4.39 (0.76)
300	5.52 (0.58)	5.31 (0.61)	4.47 (0.86)
400	5.47 (0.49)	5.30 (0.45)	4.83 (0.64)
Overall	5.42 (0.86)	5.25 (0.55)	4.56 (0.86)

trustworthy than faces paired with negative behaviors (M = 4.56, SD = 0.86), $t(79) = 8.00$, $p < .001$, $d = .98$.

Face recognition accuracy was calculated for each participant using A', a nonparametric measure of sensitivity that requires fewer assumptions than the conventional, parametric measure d' and is more robust to violations (Stanislaw & Todorov, 1999). The mean A' (M = 0.80, SD = 0.13) was significantly higher than the chance level of 50%, $t(79) = 20.38$, $p < .001$. Recognition did not vary across group-size conditions, $F < 1$. Finally, recognition accuracy was positively correlated with the difference between trustworthiness ratings of faces paired with positive and negative behaviors, $r(80) = .31$, $p < .005$.

Participants were able to form and retrieve trustworthiness associations with faces. Surprisingly, this effect did not vary as a function of the number of faces and behaviors presented to participants. We also found that the ability to recognize the presented faces was correlated with the ability to retrieve trustworthiness associations upon the face presentation, although the correlation was not strong.

EXPERIMENT 2

The objective of Experiment 2 was to replicate the findings of Experiment 1 with a better control. Specifically, in Experiment 1, the filler place photos were neither unique nor paired with unique descriptions, and therefore may have created a possible confound in that participants who viewed fewer faces were required to form fewer impressions than participants who viewed more faces. For Experiment 2, new place photos were selected so that all photos showed unique places and the place photos were paired with unique descriptive statements. The refining of the place stimuli also allowed us to examine whether the ability to retrieve valence information was unique to faces, or could generalize to place stimuli as well.

METHODS

Participants. Eighty members of the Princeton community (25 males, mean age 20.23 years) participated in exchange for 1 hour of course credit or $12.

Materials. The 400 photos of faces and their accompanying behaviors selected in Experiment 1 were used again. A new subset of 400 place photos was selected from the Konkle et al. (2010) study so that all photos displayed unique places. Five hundred statements describing places were generated in our lab. Examples of negative, neutral, and positive place descriptions include "this pond is riddled with parasites," "four types of algae grow here," and "thousands of bright stars can be seen here at night," respectively. A small validation study was run to obtain valence ratings on these descriptions. Eighteen Princeton University undergraduate students (8 males, mean age 19.39 years) participated for course credit. Participants were asked to read each of the 500 statements describing places and rate how negative or positive each statement was on a scale of 1 to 9, where 1 meant very negative, 5 meant neutral, and 9 meant very positive. Statements were presented in random order. Participants were given unlimited time to make each judgment, but were instructed to report their first impression of the statements and not to spend too much time on any one statement. Out of the 500 statements, the 132 most positive statements (which received ratings ranging from 6.11 to 7.89), 136 most neutral statements (4.56 to 5.94), and 132 most negative statements (1.67 to 3.72) were selected for use in this experiment. As with photos of faces, each place photo was paired with only one description during the experiment.

Procedure. The impression formation task was nearly identical to the one described in Experiment 1. However, participants were instructed that they would be viewing photos of places paired with descriptions in addition to viewing faces, and in addition to forming impressions of faces, participants were instructed to form impressions of all of the places viewed by imagining the places portraying their descriptions. Each participant was randomly assigned to one of four conditions: 100 faces and 400 places, 200 faces and 300 places, 300 faces and 200 places, or 400 faces and 100 places. Thus, every participant viewed 500 photos total in a 4 (group-size condition) by 3 (valence) by 2 (faces vs. places) mixed factorial design.

The ratings task was very similar to that used in Experiment 1. Here, 60 place photos were rated in addition to the 60 face photos, and both types of stimuli were rated on valence instead of trustworthiness. Specifically, participants were asked, "How good or bad is this person?" or "How good or bad is this place?" as was appropriate. Ratings were made on a scale of 1 to 9, where 1 meant very bad, 5 meant neutral, and 9 meant very good. Like the face photos, the same 60 place photos were rated by all participants. The rated place photos were weighted equally for valence (20 negative, 20 neutral, and 20 positive) and were presented in the middle 300 photos during the impression-formation task to avoid primacy and recency effects.

The recognition task was identical to that used in Experiment 1 with the addition of 20 place photos, 10 of which had been shown during the impression formation task but not rated in the valence-ratings task and 10 of which had not been shown previously in the experiment.

RESULTS AND DISCUSSION

Valence Ratings Task. Participants rated both faces and places paired with positive descriptions more positively than faces and places paired with negative de-

TABLE 2. Means (SD) of Goodness Judgments of Faces and Places as a Function of the Valence of the Descriptions Associated with Them and the Number of Faces/Places-and-Descriptions Pairs

	Faces			Places		
Number	Positive	Neutral	Negative	Positive	Neutral	Negative
100	5.33 (0.81)	5.21 (0.85)	4.49 (0.76)	5.75 (0.61)	5.49 (0.45)	4.70 (0.76)
200	5.43 (0.63)	5.41 (0.39)	4.68 (0.65)	5.75 (0.74)	5.36 (0.55)	4.95 (0.74)
300	5.68 (0.48)	5.44 (0.34)	4.80 (0.50)	5.82 (0.58)	5.50 (0.46)	4.84 (0.69)
400	5.21 (0.66)	5.41 (0.48)	4.81 (0.51)	5.95 (0.98)	5.55 (0.88)	4.93 (0.78)
Overall	5.41 (0.67)	5.37 (0.55)	4.69 (0.62)	5.82 (0.74)	5.48 (0.60)	4.85 (0.74)

scriptions (see Table 2), as revealed by a large effect of valence, $F(2, 152) = 127.24$, $p < .001$, $\eta^2 = .63$. Participants' ratings were slightly more positive for places (M = 5.38, SD = 0.58) than for faces (M = 5.16, SD = 0.49), $F(1, 76) = 7.30$, $p < .009$, $\eta^2 = .09$.

Because the interaction of these two main effects was significant, $F(2, 152) = 5.36$, $p < .006$, $\eta^2 = .07$, we conducted separate analyses for places and faces. There were no other significant effects in the overall analysis, Fs < 1.5, ps > .22.

Replicating Experiment 1, participants rated faces paired with positive and neutral behaviors more positively than faces paired with negative behaviors, as revealed by a significant main effect of valence, $F(2, 152) = 64.80$, $p < .001$, $\eta^2 = .46$. This effect did not vary across group-size condition (Fs < 1.5, ps > .18 for the main effect of group-size condition and the interaction). Post-hoc comparisons (Bonferroni correction applied) showed that faces paired with neutral behaviors (M = 5.37, SD = 0.55) were rated significantly more positively than faces paired with negative behaviors (M = 4.69, SD = 0.62), $t(79) = 10.11$, $p < .001$, $d = 1.14$, whereas faces paired with positive behaviors (M = 5.41, SD = 0.67) were rated no more positively than faces paired with neutral behaviors, this difference was small and did not reach significance, $t(79) < 1$.

Participants also rated place photos paired with positive and neutral descriptions as more positive than places paired with negative descriptions, as shown by a significant main effect of valence, $F(2, 152) = 84.04$, $p < .001$, $\eta^2 = .53$. Participants rated place photos paired with positive descriptions (M = 5.82, SD = 0.74) more positively than those paired with neutral descriptions (M = 5.48, SD = 0.60), $t(79) = 6.60$, $p < .001$, $d = .74$; participants also rated places paired with neutral descriptions more positively than those paired with negative descriptions (M = 4.85, SD = 0.74), $t(79) = 8.68$, $p < .001$, $d = .98$ (Bonferroni correction applied for both comparisons). Like for faces, the effect of valence did not vary across group-size conditions (Fs < 1 for the main effect of group-size condition and the interaction).

Recognition Task. As in Experiment 1, A' scores for sensitivity were calculated based on the participants' performances on the face recognition task, and here we also calculated A' scores for place recognition. One participant in the 200 faces condition and two participants in the 400 places condition neither correctly recognized any of the items as shown during the experiment nor falsely recognized any novel items as having been shown, and so we were unable to calculate A' scores for those participants. As a result, they could not be included in the recognition task analysis.

Replicating Experiment 1, the mean *A′* for faces ($M = 0.76$, $SD = 0.14$) was significantly higher than the chance level of 50% (i.e., *A′* of 0.5), $t(78) = 16.09$, $p < .001$. The mean *A′* for places ($M = 0.74$, $SD = 0.14$) was also significantly higher than chance, $t(77) = 15.33$, $p < .001$. The recognition scores did not vary across the group-size conditions, $F(3, 75) = 1.82$, $p = .15$, for faces and $F(3, 74) = 1.06$, $p = .37$, for places. The *A′* scores for faces and places were not significantly correlated, $r(77) = .16$, $p = .16$. In contrast to Experiment 1, we found no relationship between individual participants' *A′* scores and the difference in their valence ratings between faces paired with positive and negative behaviors, $r(79) = .05$, $p = .64$. For places, there was a positive relationship between *A′* scores and the difference between ratings of place photos paired with positive and negative descriptions, $r(78) = .24$, $p < .04$.

Replicating Experiment 1, participants were able to retrieve affective associations with faces even after being presented with 400 face and behavior pairs. Again, this effect did not vary as a number of face-behavior pairs. Unlike in Experiment 1, there was no significant difference between ratings of faces paired with positive behaviors and faces paired with neutral behaviors. Finally, face recognition accuracy was not correlated with the difference in judgments of faces paired with positive and faces paired with negative behaviors. We found very similar effects for places as for faces; if anything, the effects were slightly stronger for places than for faces.

EXPERIMENT 3

Experiment 3 had two main objectives. The first was to probe further the limit of the number of affective associations that participants could form and retrieve by including a 500 faces and behaviors condition. The second was to test whether the effects generalize to a different behavioral domain. In the first two experiments, most of the behaviors can be classified as related to morality. This could explain the negativity bias found in these experiments for valence judgments of faces, with a stronger difference between faces paired with negative behaviors compared to neutral behaviors than between faces paired with positive compared to neutral behaviors. This bias is well documented for behaviors that fall under the morality domain (Skowronski & Carlston, 1989; Wojciszke, 2005).

In this experiment, we included two types of behaviors: morality- and competence-related. Given the greater importance of morality information (Wojciszke, 2005), we expected that the associations with faces would be stronger for morality-related than for competence-related behaviors. The difference between morality- and competence-related information is also interesting, because negative information is considered more diagnostic than positive information in the morality domain (Crocker, Fiske, & Taylor, 1984; Reeder & Coovert, 1986; Reeder & Spores, 1983), while positive information is considered more diagnostic in the competence domain (Reeder & Fulks, 1980; Reeder, Messick, & Van Avermaet, 1977; Skowronski & Carlston, 1987; Wojciszke, 2005). Because of these differences in diagnosticity, we anticipated to find a positivity bias for competence-related behaviors rather than a negativity bias.

Finally, to increase statistical power, we analyzed the data from all three experiments to look for evidence that the strength of the associations between faces and behaviors decreases as the number of faces and behaviors increases.

METHODS

Participants. Eighty-three members of the Princeton community (38 males, mean age 20.23 years) participated in exchange for a payment of $12.

Materials. To ensure that the effects observed in Experiments 1 and 2 generalize across faces, a new set of 250 unique male and 250 unique female faces was randomly selected from the FERET Grayscale Database CDs 1 and 2. All selected pictures were forward facing, had adequate lighting, were un-posed, and did not display emotional expressions.

Three hundred sixty-three statements from Fuhrman et al. (1989) were selected for use in this experiment. These statements were selected because they fell into clear classifications as either positive or negative on either "goodness" or "intelligence," or neutral in both domains, according to the cluster analysis presented by those authors. Of the statements selected from Fuhrman et al. (1989), the positive "goodness" statements had means ranging from 6.95 to 9.08 and the negative "goodness" means ranged from 0.51 to 4. The positive "intelligence" statements had mean ratings from 6.92 to 9.77 and the negative "intelligence" from 1 to 4.49. The neutral statements had mean "goodness" ratings ranging from 4.74 to 6.92 and mean "intelligence" ratings from 4.6 to 7.09. One hundred thirty-seven statements were generated in our lab so that the final experiment contained 500 behaviors total. The final set of behaviors consisted of 83 negative goodness, 83 positive goodness, 83 negative intelligence, 83 positive intelligence, and 168 neutral behaviors. Faces were randomly assigned to unique behaviors. Only faces paired with behaviors that were rated in Fuhrman et al. (1989) were rated in the trait-ratings task described below. Examples of negative and positive competence statements include "she does not know the multiplication tables," and "she can speak four languages." For the faces rated in the valence ratings task, three counterbalanced versions were created in each group-size condition so that for different participants, the judged faces rotated among being paired with negative, neutral, and positive behaviors within their assigned domain during the impression-formation task. The aim of this counterbalancing was to avoid any artifacts that may have arisen from specific face-behavior pairings. The same place photos and descriptions used in Experiment 2 were used again here.

Procedure. The impression formation task was identical to that in Experiment 2, except that the 100 faces and 400 places condition was replaced by a 500 faces condition, which did not include any place photos. Each participant was randomly assigned to one of four conditions: 200 faces and 300 places, 300 faces and 200 places, 400 faces and 100 places, or 500 faces. Thus, every participant viewed 500 photos

TABLE 3. Means (SD) of Judgments of Faces as a Function of the Valence of the Behaviors Associated with the Faces and the Number of Faces-and-Behaviors Pairs

	Morality-related behaviors			Competence-related behaviors		
Number	Positive	Neutral	Negative	Positive	Neutral	Negative
200	5.43 (0.79)	5.17 (0.73)	4.97 (0.83)	5.75 (0.81)	5.43 (0.76)	5.12 (0.66)
300	5.27 (0.71)	5.05 (0.77)	4.91 (0.81)	5.51 (0.78)	5.29 (0.64)	4.90 (0.73)
400	5.30 (0.99)	5.14 (0.85)	4.67 (0.72)	5.59 (0.57)	5.24 (0.62)	4.98 (0.81)
500	5.44 (0.72)	5.42 (0.65)	4.95 (0.72)	5.67 (0.67)	5.34 (0.54)	5.20 (0.63)
Overall	5.36 (0.80)	5.20 (0.75)	4.90 (0.77)	5.63 (0.71)	5.33 (0.64)	5.05 (0.71)

Note. Participants made trustworthiness judgments of faces associated with morality-related behaviors and competence judgments of faces associated with competence-related behaviors.

total in a 4 (group-size condition) by 3 (valence) by 2 (behavioral domain: morality vs. competence) mixed factorial design.

The trait-ratings task was very similar to the trustworthiness ratings task used in Experiment 1; because the 500 faces condition did not contain any place photos, places were not rated in this experiment. Here, faces that had been paired with positive or negative morality behaviors during the impression-formation task were rated on trustworthiness and faces paired with positive or negative intelligence behaviors were rated on competence. Ratings were made on a scale of 1 to 9, where 1 meant not at all trustworthy or not at all competent and 9 meant very trustworthy or very competent, as was appropriate. As in the previous experiments, all participants rated the same faces. However, here, 48 photos were rated on trustworthiness and 48 photos were rated on competence, resulting in 96 total trait ratings. The recognition task was the same as the one used in Experiment 1.

RESULTS

Ratings Task. Participants rated faces paired with positive and neutral behaviors more positively than faces paired with negative behaviors (see Table 3), as confirmed by a significant main effect of valence, $F(2, 158) = 41.25$, $p < .001$, $\eta^2 = .34$. The only other significant effect in the overall analysis was the effect of type of behavior, $F(1, 79) = 13.40$, $p < .001$, $\eta^2 = .15$ (Fs < 1.02 for all other effects). Participants rated faces paired with competence-related behaviors ($M = 5.34$, $SD = 0.57$) more positively than faces paired with morality-related behaviors ($M = 5.15$, $SD = 0.64$). Interestingly, the interaction of valence by type of behaviors did not reach significance ($p = .36$), indicating that participants rated faces matched with positive, negative, or neutral statements to a similar extent when statements referred to competence and to trustworthiness. An analysis of the simple effects (Bonferroni correction applied) showed that faces paired with positive behaviors ($M = 5.50$, $SD = 0.65$) were rated more positively than faces paired with neutral behaviors ($M = 5.26$, $SD = 0.58$), $t(82) = 4.33$, $p < .001$, $d = .48$. In turn, the latter were rated more positively than faces paired with negative behaviors ($M = 4.96$, $SD = 0.68$), $t(82) = 6.03$, $p < .001$, $d = .66$.

Although the interaction of valence and type of behavior did not reach significance, inspection of Table 3 (last row) shows that the pattern of means for trustworthiness and competence judgments was consistent with the expected asymmetry of positive and negative information. Whereas trustworthiness judgments exhibited a negativity bias, competence judgments exhibited a weak positivity bias. To explore these effects further, we conducted separate analyses for trustworthiness and competence judgments.

For trustworthiness judgments, the effect of valence was significant, $F(2, 158) = 18.02$, $p < .001$, $\eta^2 = .19$. There were no other significant effects. Post hoc tests (Bonferroni correction applied) showed that that faces paired with positive behaviors were rated as more trustworthy than faces paired with neutral behaviors, $t(82) = 2.12$, $d = .23$, although the effect did not reach the corrected significance level. In turn, the latter were rated as more trustworthy than faces paired with negative behaviors, $t(82) = 4.12$, $p < .001$, $d = .45$.

For competence judgments, the effect of valence was also significant and larger in size than the effect for trustworthiness judgments, $F(2, 158) = 31.90$, $p < .001$, $\eta^2 = .29$ (there were no other significant effects). This was largely driven by the difference in judgments of faces associated with positive behaviors and faces associated with neutral behaviors, $t(82) = 4.07$, $p < .001$, $d = .45$, an effect which was larger than the effect for trustworthiness judgments. The difference between competence judgments of faces associated with negative behaviors and faces associated with positive behaviors, $t(82) = 4.53$, $p < .001$, $d = .50$, was similar in size to the corresponding difference for trustworthiness judgments.

Recognition Task. Recognition accuracy was calculated as in Experiments 1 and 2. We were unable to calculate A' scores for two participants who neither correctly recognized any of the items as shown during the experiment nor falsely recognized any novel items as having been shown; these participants were not included in the recognition task analysis. The mean A' ($M = 0.71$, $SD = 0.12$) was significantly higher than the chance level of 50%, $t(80) = 16.17$, $p < .001$. The recognition scores did not vary across group-size conditions, $F < 1$. The A' scores were positively correlated with the differences between ratings of faces paired with positive and negative behaviors, $r(81) = .22$, $p < .05$.

Replicating the first two experiments, participants retrieved affective associations with faces and this effect did not vary with the number of faces and behaviors. The effects were similar for morality-related and competence-related behaviors. Although the means were in the predicted direction of the expected positivity-negativity asymmetry for competence and trustworthiness judgments, respectively, the interaction of valence and type of behavior did not reach significance. It may be that this asymmetry is more pronounced in consecutive judgments of the same individuals (cf. Mende-Siedlecki, Baron, & Todorov, 2013).

Analyses across All Experiments. In three experiments, we failed to find evidence that the strength of valence associations between faces and behaviors decreases as the number of faces and behaviors increases (see Figure 1A). A likely reason is that the experiments were statistically underpowered to detect the interaction of valence and group-size. For this reason, we conducted an analysis across the experiments. The valence judgments of faces were submitted to 3 (Valence) × 3 (Experiment) × 4 (Group-size condition) mixed-subjects ANOVA. In this analysis, we treated the four levels of group-size in Experiment 3 the same as the four levels

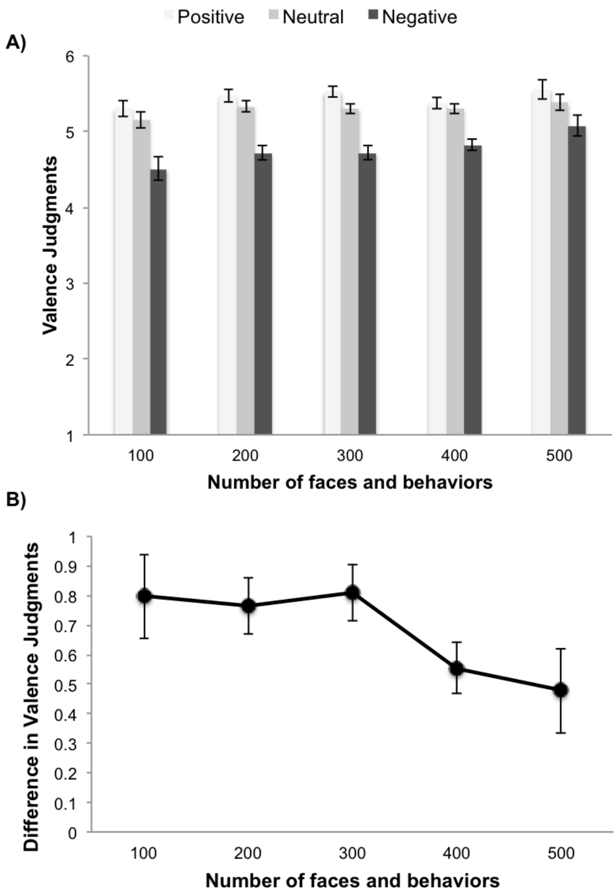

FIGURE 1. Analysis of valence judgments of faces across all experiments. (A) Valence judgments of faces as a function of the valence of the behaviors associated with them and the number of face-behavior pairs. (B) Differences in valence judgments between faces paired with positive and faces paired with negative behaviors as a function of the number of face-behavior pairs. The data for the 100 face-behavior pairs condition includes Experiments 1 and 2; the data for the 200, 300, and 400 pairs conditions includes Experiments 1, 2, and 3; and the data for the 500 pairs condition includes only Experiment 3. Standard errors are standard errors of the mean.

in the other two experiments despite the differences in numbers of faces and behaviors (an alternative analysis, focusing only on the shared conditions across the three experiments—200, 300, and 400 faces and behaviors—produced very similar results). In addition to the significant effect of valence, $F(2, 462) = 164.29$, $p < .001$, $\eta^2 = .42$, there was a small significant interaction effect of experiment and valence, $F(4, 462) = 5.69$, $p < .001$, $\eta^2 = .05$. This interaction was driven by the larger differences between judgments of faces paired with positive behaviors and faces paired with neutral behaviors in Experiment 3 than in Experiments 1 and 2. There were no other significant effects in the analysis. The interaction of valence and group-size remained small in size despite the large sample size, $F(6, 462) = 1.75$, $p = .11$, $\eta^2 = .022$.

In a final test, we computed a linear contrast on the differences between judgments of faces paired with positive behaviors and faces paired with negative behaviors as the faces and behaviors increased from 100 to 500. This contrast showed a small effect, $t(238) = 1.98$, $p = .049$, $d = .13$, suggesting that the strength of valence associations decreases as the number of faces and behaviors increases. As shown in Figure 1B, the pattern of results suggests that the strength of the associations remains the same for 100 to 300 faces and behaviors, and then gradually decreases with the increase in faces and behaviors. However, given the post hoc nature of this analysis on data from different experiments with different measures of valence judgments, this result should be interpreted with caution.

Because there were some inconsistencies in the correlations between face recognition accuracy and valence judgments across the experiments, we correlated the A' scores and the differences between ratings of faces paired with positive and negative behaviors combining the three samples. This correlation was significant, $r(240) = .20$, $p < .001$. Using an alternative measure—the difference between the hits and the false alarms—which uses data from all participants and was highly correlated with A' ($r = .85$) led to the same result, $r(243) = .29$, $p < .001$. Thus, across experiments there was a modest correlation between the ability to recognize faces and retrieve affective associations about these faces.

GENERAL DISCUSSION

Consistent with prior studies (Bliss-Moreau et al., 2008; Cassidy & Gutchess, 2014; Todorov & Uleman, 2002, 2003), we found that participants are able to form and retrieve trustworthiness associations about faces after learning a single behavioral act. Despite the fact that participants sometimes described the impression-formation task as "intense" or "tiring," they were nevertheless able to form and retrieve trustworthiness associations even when viewing 500 face-behavior pairs. In fact, across all three experiments we failed to find evidence that the strength of the associations decreased with the increase in number of face-behavior pairs. The only evidence we found for a possible decrease in the strength of the associations was based on a post-hoc analysis combining data from all the experiments. This analysis suggested that after 300 face-behavior pairs, there may be a gradual decrease in the ability to retrieve trustworthiness associations with faces. Given the post-hoc nature of this analysis, this finding should be interpreted with caution. In any case, this potential decline is informative for future studies testing the limits of forming trustworthiness associations with faces.

An unexpected result that emerged from Experiment 2 was the robustness with which participants were able to recall valence information associated with places. Similarly to faces, participants were able to retrieve valence information about places even when viewing photos and descriptions of 400 places. However, ongoing work in our lab suggests that the processes of forming affective associations with faces and places may be different. Using a paradigm very similar to the present one, we find that associations with faces are formed only after learning person-specific information. In contrast, associations with places can be formed after learning both place-specific and person-specific information.

Another facet of our findings is the pattern of valence ratings of faces paired with different types of affective behaviors. Specifically, we consistently found that the difference between trustworthiness judgments of faces paired with negative behaviors, and faces paired with neutral ones, was larger than the difference between judgments of the latter and judgments of faces paired with positive behaviors. This result is consistent with the literature on negativity bias in trait inferences (Carlston & Skowronski, 2005; Shimizu, 2012; Skowronski & Carlston, 1989; Wells, Skowronski, Crawford, Scherer, & Carlston, 2011), which shows that negative behaviors produce larger inference effects than positive ones, are associated with more inference certainty and extremity, and are recalled more accurately than positive behaviors when learned only once. An alternative but not mutually exclusive explanation for the negativity bias is that the failure to detect morally bad people may be more costly than the failure to detect morally good people.

Furthermore, the behaviors tested in the first two experiments were morality-related behaviors. Much literature has demonstrated that negativity biases are much more pronounced for such behaviors; and may be absent or even reversed for other behavioral domains. Specifically, for behaviors that connote morality, negative behaviors are weighed more heavily in judgments than positive ones (Crocker et al., 1984; Reeder & Coovert, 1986; Reeder & Spores, 1983). In contrast, for behaviors that connote competence or ability, positive behaviors are weighted more heavily than negative ones (Reeder & Fulks, 1980; Reeder et al., 1977; Skowronski & Carlston, 1987; Wojciszke, 2005). However, although the means were in the predicted direction of the expected positivity-negativity asymmetry, we found similar effects for morality-related and competence-related behaviors. It may be that such biases are most pronounced when people have to make consecutive judgments of the same individuals (Mende-Siedlecki et al., 2013) rather than when judging numerous different individuals. This finding certainly warrants additional research.

Across the experiments, there was a modest correlation between the ability to recognize faces and the ability to judge them according to the valence implications of the associated behaviors. However, it should be noted that the face recognition test was not very powerful as it included a very small set of faces and these faces did not overlap with the learned faces. Because the designs of the current studies were already exhausting for participants, we were unable to test their memories for the exact behaviors viewed with the faces. However, it is plausible that the affective associations may be completely dissociated from the memory for the specific behaviors on which these associations are based. Past studies have demonstrated that people tend to have poor memory for specific behaviors (Todorov & Uleman, 2002), although these studies used only 120 behaviors, and that this memory doesn't predict the trait associations with the faces. This finding is consistent with work by Tranel and Damasio (1993) who studied a densely amnesiac patient with large lesions in cortical and subcortical limbic structures, the temporal pole, the insular cortices, and other regions. Even with substantial portions of his brain severely damaged, the patient reliably preferred the faces of caregivers who had treated him nicely over others despite the fact that he did not recognize their faces explicitly or implicitly. In an earlier study, Johnson and colleagues examined patients with Korsakoff's syndrome, a condition characterized by anterograde and retrograde amnesia and severe memory loss (Johnson et al., 1985). The patients first read about two characters—one bad and one good—and while learning this information were presented with pictures of the characters. Subsequently, they preferred the face of the "good" character over the "bad" one despite being unable to voluntarily access the information they learned about the characters. Finally, Todorov and Olson (2008) observed that a patient with amnesia due to lesions in the hippocampus performed just as well as control participants at learning trustworthy and untrustworthy associations with faces after being presented with faces and positive or negative behaviors. Therefore, based on past studies and the current analyses, there is not strong enough evidence to conclude that the ability to recognize specific faces is necessary for recalling trustworthiness information about faces. Rather, further studies are needed to examine the nature and strength of such a relationship.

Another future direction of this research might more closely examine the relationship between the upper limit of affective inferences that can be retrieved and social network size. Though our participants were able to retrieve many more affective associations than predicted by the social network studies mentioned above, more recent studies, using a variety of survey and statistical methods, have estimated the average social network size in the United States to be much larger than predicted by Dunbar's social brain hypothesis (McCarty, Killworth, Bernard, Johnsen, & Shelley, 2001; McCormick, Salganik, & Zheng, 2010; Zheng, Salganik, & Gelman, 2006). Moreover, detailed studies of hunter-gatherers suggest that each individual may know more than 1000 other individuals despite the fact that they live in small residential bands ranging from 20 to 50 individuals (Hill, Wood, Baggio, Hurtado, & Boyd, 2014). Larger estimates of social network size are more consistent with our results that people can form and retrieve as many as 500 affective associations, but it is still unclear whether, and to what degree, these two features are correlated with each other.

Learning about the trustworthiness of other people is a surprisingly robust process. Our findings suggest that this learning is based on global affect-based inferences. This parallels findings from research on judgments from facial appearance, where trustworthiness judgments closely approximate global valence inferences (Oosterhof & Todorov, 2008; Sutherland et al., 2013; Todorov et al., 2008). These inferences are not only efficient but also emerge early in development (Cogsdill, Todorov, Spelke, & Banaji, 2014). An important question for future research is to what extent and under what conditions trustworthiness judgments can be dissociated from affect-based judgments.

REFERENCES

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

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

Carlston, D. E., & Skowronski, J. J. (2005). Linking versus thinking: Evidence for the different associative and attributional bases of spontaneous trait transference and spontaneous trait inference. *Journal of Personality and Social Psychology, 89*, 884-898.

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

Cassidy, B. S., & Gutchess, A. H. (2014). Influences of appearance-behavior congruity on memory and social judgements. *Memory*. E-pub ahead of print (Sep 2, 1-17).

Chang, L. J., Doll, B. B., van't Wout, M., Frank, M. J., & Sanfey, A. G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology, 61*(2), 87-105.

Cogsdill, E., Todorov, A., Spelke, E., & Banaji, M. R. (2014). Inferring character from faces: A developmental study. *Psychological Science, 25*, 1132-1139.

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

Crocker, J., Fiske, S., & Taylor, S. (1984). Schematic bases of belief change. In J. Eiser (Ed.), *Attitudinal judgment* (pp. 197-226). New York: Springer.

Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution, 22*, 469-493.

Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences, 16*, 681-735.

Dunbar, R. I. M. (1998). The social brain hypothesis and its implications for social evolution. *Annals of Human Biology, 36*, 562-572.

Dunbar, R. I. M., Duncan, N. D. C., & Marriott, A. (1997). Human conversational behavior. *Human Nature, 8*, 231-246.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*, 598-601.

Ewing, L., Caulfield, F., Read, A., & Rhodes, G. (2014). Perceived trustworthiness of faces drives trust behaviour in children. *Developmental Science, 18*, 327-334.

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, and Computers, 21*, 587-597.

Goodwin, G. P., Piazza, J., & Rozin, P. (2013). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology, 106*, 1-21.

Johnson, M. K., Kim, K. J., & Risse, G. (1985). Do alcoholic Korsakoff's syndrome patients acquire affective reactions? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 22-36.

Hill, K. R., Wood, B. M., Baggio, J., Hurtado, A. M., & Boyd, R. T. (2014). Hunter-gatherer inter-band interaction rates: Implications for cumulative culture. *PLoS One, 9*(7), e102806. doi: 10.1371/journal.pone.0102806

Killworth, P. D., Bernard, H. R., & McCarty, C. (1984). Measuring patterns of acquaintanceship. *Current Anthropology, 25*, 381-397.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science, 21*, 1551-1556.

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say "Hello"? Personality impressions from brief novel voices. *PLoS ONE, 9*(3), e90779. doi: 10.1371/journal.pone.0090779

McCarty, C., Killworth, P. D., Bernard, H. R., Johnsen, E. C., & Shelley, G. A. (2001). Comparing two methods for estimating network size. *Human Organization, 60*, 28-39.

McCormick, T. H, Salganik, M. J., & Zheng, T. (2010). How many people do you know?: Efficiently estimating personal network size. *Journal of the American Statistical Association, 105*, 59-70.

Mende-Siedlecki, P., Baron, S. G., & Todorov, A. (2013). Diagnostic value underlies asymmetric updating of impressions in the morality and ability domains. *Journal of Neuroscience, 33*, 19406-19415.

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

Phillips, P. J., Moon, H., Rizvi, S. A., & Rauss, P. J. (1999). *The FERET evaluation methodology for face-recognition algorithms* (National Institute for Standards and Technology Internal Report No. 6264). Retrieved from http://www.itl.nist.gov/iad/humanid/feret/doc/feret_methodology_nist_ir_6264.pdf.

Reeder, G. D., & Coovert, M. D. (1986). Revising an impression of morality. *Social Cognition, 4*, 1-17.

Reeder, G. D., & Fulks, J. L. (1980). When actions speak louder than words: Implicational schemata and the attribution of ability. *Journal of Experimental Social Psychology, 16*, 33-46.

Reeder, G. D., Messick, D. M., & Van Avermaet, E. (1977). Dimensional asymmetry in attributional inference. *Journal of Experimental Social Psychology, 13*, 46-57.

Reeder, G. D., & Spores, J. M. (1983). The attribution of morality. *Journal of Personality and Social Psychology, 44*, 736-745.

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PLoS ONE, 7*(3), e34293.

Schliht, E. J., Shinsuke, S., Camerer, C. F., Battaglia, P., & Nakayama, K. (2010). Human wagering behavior depends on opponents' faces. *PLoS ONE, 5*, e11663.

Shimizu, Y. (2012). Spontaneous trait inferences among Japanese children and adults: A developmental approach. *Asian Journal of Social Psychology, 15*, 112-121.

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personal Social Psychology, 52*, 689-699.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin, 105*, 131-142.

Standing, L. (1973). Learning 10000 pictures. *The Quarterly Journal of Experimental Psychology, 25*, 207-222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science, 19*, 73-74.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers, 31*, 137-149.

Stirrat, M., & Perrett, D. (2010). Valid facial cues to cooperation and trust. *Psychological Science, 21*(3), 349-354.

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition, 127*, 105-118.

Tingley, D. (2014). Face-off: Facial features and strategic choice. *Political Psychology, 35*, 35-55.

Todorov, A., Loehr, V., & Oosterhof, N. N. (2010). The obligatory nature of holistic processing of faces in social judgments. *Perception, 39*, 514-532.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology, 66*, 519-545.

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., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*, 813-833.

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

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology, 83*, 1051-1065.

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

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

Tranel, D., & Damasio, A. R. (1993). The covert learning of affective valence does not require structures in hippocampal system or amygdala. *Journal of Cognitive Neuroscience, 5*, 79-88.

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science, 10*, 72-81.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. *Advances in Experimental Social Psychology, 28*, 211-279.

van't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision making. *Cognition, 108*, 796-803.

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. (2013). When physical similarity matters: Mechanisms underlying affective learning generalization to the evaluation of novel faces. *Journal of Experimental Social Psychology, 49*, 661-669.

Wells, B. M., Skowronski, J. J., Crawford, M.T., Scherer, C. R., & Carlston, D. E. (2011). Inference making and linking both require thinking: Spontaneous trait inference and spontaneous trait transference both rely on working memory capacity. *Journal of Experimental Social Psychology, 47*, 1116-1126.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after 100 ms exposure to a face. *Psychological Science, 17*, 592-598.

Wojciszke, B. (2005). Morality and competence in person and self-perception. *European Review of Social Psychology, 16*, 155-188.

Zheng, T., Salganik, M. J., & Gelman, A. (2006). How many people do you know in prison? *Journal of the American Statistical Association, 101*, 409-423.