# Do We Spontaneously Form Stable Trustworthiness Impressions From Facial Appearance?

André Klapper
Radboud University

Ron Dotsch
Utrecht University and Radboud University

Iris van Rooij and Daniël H. J. Wigboldus
Radboud University

It is widely assumed among psychologists that people spontaneously form trustworthiness impressions of newly encountered people from their facial appearance. However, most existing studies directly or indirectly induced an impression formation goal, which means that the existing empirical support for spontaneous facial trustworthiness impressions remains insufficient. In particular, it remains an open question whether trustworthiness from facial appearance is encoded in memory. Using the 'who said what' paradigm, we indirectly measured to what extent people encoded the trustworthiness of observed faces. The results of 4 studies demonstrated a reliable tendency toward trustworthiness encoding. This was shown under conditions of varying context-relevance, and salience of trustworthiness. Moreover, evidence for this tendency was obtained using both (experimentally controlled) artificial and (naturalistic varying) real faces. Taken together, these results suggest that there is a spontaneous tendency to form relatively stable trustworthiness impressions from facial appearance, which is relatively independent of the context. As such, our results further underline how widespread influences of facial trustworthiness may be in our everyday life.

*Keywords:* face perception, spontaneity, trait inferences, trustworthiness, "who said what" paradigm



It is widely assumed among psychologists that people spontaneously form trustworthiness impressions of newly encountered people from their facial appearance (Marzi, Righi, Ottonello, Cincotta, & Viggiano, 2014; Todorov, Said, Engell, & Oosterhof, 2008). These face-based impressions have been shown to influence important outcomes such as investment decisions in a trust game (Chang, Doll, van 't Wout, Frank, & Sanfey, 2010; Rezlescu, Duchaine, Olivola, & Chater, 2012; Schlicht, Shimojo, Camerer, Battaglia, & Nakayama, 2010; Stirrat & Perrett, 2010; van 't Wout & Sanfey, 2008) and sentencing decisions (Porter, ten Brinke, & Gustaw, 2010). If these impressions are truly formed spontaneously, the impact of facial appearance on our behavior would not only be profound but also frequent in our daily life.

The assumption that face-based trustworthiness impressions are formed spontaneously originated mainly from four lines of evidence. First, it was shown that people are able to infer trustworthiness even from minimal exposure to faces (Willis & Todorov, 2006). Second, it was shown that people judge faces mostly on trustworthiness when asked to form impressions of displayed faces (Todorov et al., 2008). However, in both cases participants were explicitly instructed to form an impression. Therefore, it does not follow from these findings that trustworthiness inferences occurred spontaneously: that is, without an impression formation instruction.

Third, it was shown that people are influenced by the facial trustworthiness of a player in a trust game (Chang et al., 2010; Rezlescu et al., 2012; Schlicht et al., 2010; Stirrat & Perrett, 2010; van 't Wout & Sanfey, 2008). However, given that performance in a trust game depends on how well the trustworthiness of the other player is judged, asking a participant to play a trust game can be seen as an implicit instruction to judge trustworthiness.

Fourth, neurophysiological responses that may potentially reflect trustworthiness inferences have been shown to occur even in a task in which the trustworthiness of displayed faces is irrelevant (Engell, Haxby, & Todorov, 2007; Mende-Siedlecki, Said, & Todorov, 2013). However, the relationship between neurophysiological responses and trustworthiness inferences is not straightfor-

André Klapper, Behavioural Science Institute, Radboud University; Ron Dotsch, Department of Psychology, Utrecht University and Behavioural Science Institute, Radboud University; Iris van Rooij, Donders Institute for Brain, Cognition, and Behaviour, Radboud University; Daniël H. J. Wigboldus, Behavioural Science Institute, Radboud University.

André Klapper is now at Donders Institute for Brain, Cognition, and Behavior, Radboud University.

This research was supported by NWO Grant 464-11-036 awarded to Ron Dotsch and Daniël Wigboldus. We thank Lin Jansen for her support with the interpretation of the results. In addition, we thank Christoph Klauer for comments on a previous version of the present manuscript.

Correspondence concerning this article should be addressed to André Klapper, P.O. Box 9101, 6500 HB Nijmegen, the Netherlands. E-mail: a.klapper@psych.ru.nl



http://dx.doi.org/10.1037/pspa0000062

ward (Mende-Siedlecki et al., 2013; Said, Dotsch, & Todorov, 2010). For instance, amygdala responses have been found even when faces are varied on dimensions that have no known social meaning (Said et al., 2010; Sofer, Dotsch, Wigboldus, & Todorov, 2015). This makes it desirable to further investigate the spontaneity of trustworthiness inferences with different measures. In addition, it remains unclear whether facial trustworthiness was spontaneously encoded in memory. As such, it remains an open question whether *stable* impressions were formed spontaneously based on facial appearance.

It is worth noting that there are also various studies, which showed that people tend to spontaneously infer personality traits (including trustworthiness) from observed behaviors (Uleman, Hon, Roman, & Mokowitz, 1996; Uleman, Newman, & Moskowitz, 1996). Some of these studies have also shown that behavior-based inferences tend to become associated with the face of the target person and thus become encoded in memory (Todorov & Uleman, 2002, 2004; Van Overwalle, Drenth, & Marsman, 1999). However, none of these studies has investigated the spontaneity of trustworthiness inferences *from* facial appearance.

One may argue that the hypothesis that people spontaneously form trustworthiness impressions from facial appearance is nevertheless likely to be true for theoretical reasons. Specifically, given that detecting trustworthiness may help to cooperate with other individuals and given that detecting untrustworthiness may be vital to prevent exploitation and to promote survival, one may expect that natural selection pressures fostered the evolution of a spontaneous tendency toward face-based trustworthiness impressions. However, recent research showed that trustworthiness ratings based on facial appearance tend to be at chance level accuracy (Rule, Krendl, Ivcevic, & Ambady, 2013; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015; but see Slepian & Ames, 2016). Thus, face-based trustworthiness inferences do not appear to contain valid information about the trustworthiness of a perceived person, which suggests that trustworthiness inferences do not necessarily aid survival and reproduction. Although this does not rule out that a spontaneous tendency toward face-based trustworthiness inferences could have evolved (Zebrowitz, Fellous, Mignault, & Andreoletti, 2003), it nevertheless qualifies the theoretical plausibility that such a tendency has evolved, and that it leads to stable trustworthiness impressions.

Overall, it appears that the extent to which people spontaneously form trustworthiness impressions from facial appearance is still an open question. In particular, it remains unclear whether facial trustworthiness is spontaneously encoded in memory. Answering this question requires measuring trustworthiness encoding without mentioning trustworthiness to participants. We used the 'who said what' paradigm for this purpose (Taylor, Fiske, Etcoff, & Ruderman, 1978). This paradigm measures to what extent people rely on certain (facial or other) cues to remember the speaker of a statement. Thereby, it indirectly assesses whether a certain cue was encoded in memory. The 'who said what' paradigm has been successfully applied to investigate the spontaneous encoding of person characteristics such as sex, race, attitude, attractiveness, color of clothing, skin tone, and more (Klauer & Wegener, 1998; Maddox & Gray, 2002). Here, we apply it to indirectly measure to what extent people encode facial trustworthiness.

We tested the hypothesis that people spontaneously encode trustworthiness in four studies. We started investigating the hypothesis under conditions that may foster trustworthiness encoding and successively moved to conditions under which trustworthiness encoding would be considered more spontaneous. Specifically, Study 1 tested the hypothesis that people spontaneously encode facial trustworthiness in a context where trustworthiness was both context-relevant and made salient. Study 2 tested the same hypothesis in a trustworthiness relevant-context but without making trustworthiness additionally salient. Study 3 tested the same hypothesis in a neutral context that is more representative of a first encounter of a person. All of these studies used artificial faces that were strongly manipulated to look either trustworthy or untrustworthy. In Study 4 we investigated whether spontaneous facial trustworthiness encoding also occurs based on real faces that differ more subtly in terms of facial trustworthiness.[1] Importantly, in all studies participants were instructed to read statements without giving the instruction to form an impression of the speakers of these statements. Complementing materials (e.g., stimuli, raw data, analysis scripts, and additional results)[2] for all studies are available on the website of Open Science Framework (http://osf.io/a58zu).

## Study 1

### Method

**Participants.** Seventy-five students (54 female) of the Radboud University participated in this study ($M_{age}$ = 21.88, $SD_{age}$ = 3.15). They received five euro or partial course credit as a reward. A power analysis was not feasible for our chosen data analysis technique because it would require guessing a relatively large number of parameters (see Data Analysis section). However, we noted that several comparable 'who said what' studies found significant results with 40 or fewer participants (Klauer & Wegener, 1998).

**Procedure.** The whole study was administered in English. Participants were asked to imagine that they are about to move to another city and that they have asked eight brokers to find a house for them in return for a certain fee. This was thought to create a trustworthiness-relevant context because (a) a lot is at stake, (b) the participant is fully dependent on the broker, and (c) brokers have a motive to tell positive lies about the house. Before the main task started, participants were asked for each broker how trustworthy the face of each broker looked (1 = very *untrustworthy*, 7 = *very trustworthy*) with the face of the respective broker simultaneously displayed in the middle of the screen. This was thought to make facial (un)trustworthiness salient.

Next, the 'who said what' paradigm was used to indirectly measure trustworthiness encoding. The paradigm consisted of a

[1] To improve the readability of the paper, we do not report the studies in their chronological order. The chronological order was as follows: Study 2, Study 1, Study 3, and Study 4.

[2] We conducted three additional studies, which we do not report in this paper but which are reported in the online material. The reason for not reporting them in this paper is that we obtained insufficient model fit to interpret the results. We speculated that this happened because of problems with the employed statements in these studies, which differed from those employed in the studies reported in this paper. Importantly, the pattern of the results in all of the additional studies are in line with the results we report in this paper. Hence, the problem was not that the results were conflicting with the results in this paper but that their reliability could not be established given insufficient model fit.

learning and a test phase (Klauer & Wegener, 1998). In the learning phase, participants read statements made by four trustworthy looking and four untrustworthy looking speakers (i.e., the brokers). The individual features of the speakers were counterbalanced such that for any participant who saw the trustworthy version of a speaker, there was another participant who saw the untrustworthy version of the same speaker. Each speaker was randomly assigned to one of eight sets of statements, which each described a fictional house. Consequently, the assignment of statement sets to speakers was random across participants. The order of the statements within the set was fixed. In each learning trial, the face of the speaker was displayed in the middle of the screen. After a delay of 1500 ms, a statement (taken from the assigned statement set of the displayed speaker) appeared under the face surrounded by a speech bubble that pointed toward the face. After 8000 ms, the speaker and statement were replaced by a blank screen. The next trial started after an intertrial interval of 500 ms. The learning phase consisted of 48 trials.

In the test phase, the 48 statements from the learning phase were shown successively along with 48 distractor statements. Specifically, in each test trial a statement was shown in the middle of the screen and participants were asked whether the statement was made by one of the speakers in the learning phase ("Yes" or "No"). If "No" was answered, the next test trial was presented. If "Yes" was answered, participants were additionally asked which of the speakers had made the statement. Below this question, small pictures of the speakers were shown together with numbers that could be pressed to make the selection. The locations of the faces were randomized for each participant and the numbers were counterbalanced together with the individual features of the speakers.

After the 'who said what' task, participants were asked for each broker to indicate their willingness to pick the house that the broker recommended on a seven-point scale (1 = *not at all*, 7 = *very much*) with the face of the respective broker simultaneously displayed in the middle of the screen. Next, participants were asked in the same fashion how trustworthy each speaker looks on a seven-point scale (1 = *very untrustworthy*, 7 = *very trustworthy*). These questions served as manipulation checks of our facial trustworthiness manipulations. For both questions, the order of the brokers was randomized. Finally, participants were asked demographical questions and to what extent they had difficulty with the English language in the study ("Not at all," "Yes a little," or "Yes very much").

**Stimuli.** Pictures of trustworthy and untrustworthy looking faces were created using the FaceGen software development kit (Singular Inversions, Toronto, Canada). To manipulate trustworthiness, we used the FaceGen dimensions that were modeled by Oosterhof and Todorov (2008; Todorov, Dotsch, Porter, Oosterhof, & Falvello, 2013). Specifically, trustworthy and untrustworthy versions of eight male speakers were created through the following procedure. First, eight copies of the standard average face were morphed two standard deviations toward being male. Next, each face was morphed six standard deviations on a random dimension that was orthogonal to all known social dimensions to give each face neutral individual features (Said et al., 2010). To make the faces more realistic, each was also given an individual overlay texture that added details such as skin irregularities. In addition, we used Photoshop to give each face an individual haircut taken from pictures of real faces in the Radboud Face Database (Langner et al., 2010). Importantly, we created trustworthy and untrustworthy versions of each of the eight faces by morphing each version 2.5 standard deviations toward being trustworthy/untrustworthy (see Figure 1). In all manipulations above, skin brightness was kept constant to ensure that faces are perceived as Caucasian faces. Furthermore, 48 statements were created that described eight imaginary houses (which consisted of eight subsets of statements that described one imaginary house each) and 48 distractor statements that also described houses (without any subsets). All stimulus materials are available on Open Science Framework.

*Figure 1.* Trustworthy (left column) and untrustworthy versions (right column) of two speakers (rows). See the online article for the color version of this figure.

**Data analysis.** We used Multinomial Processing Tree (MPT) modeling to analyze the data (Riefer & Batchelder, 1988) using the 'MPTinR' package (Singmann & Kellen, 2013) in R 3.1.0 (R Core Team, 2014). This analysis strategy has been validated for the 'Who said what' paradigm and has many advantages over traditional analysis strategies (Klauer & Wegener, 1998). The employed MPT model is identical to the model used by Klauer and Wegener (1998). For ease of explanation, it is helpful to think of this model as a tree of processing stages through which participants move during the task with the most important stages being *item discrimination*, *person discrimination*, and (in this case) *(un)trustworthiness encoding* (see Figure 2).

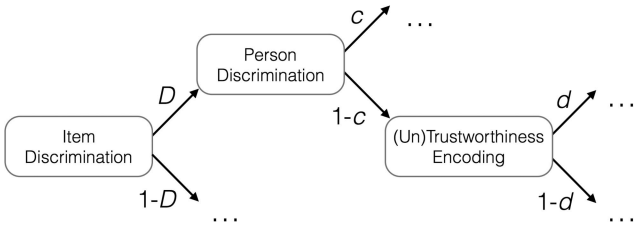


*Figure 2.* The main part of the processing tree that is assumed in the employed Multinomial Processing Tree model. *D* represents the probability of remembering the statement (item discrimination), *c* represents the probability of remembering the speaker (person discrimination), and *d* represents the probability of remembering whether the speaker was trustworthy or untrustworthy ((un)trustworthiness encoding). Success and failure probabilities add up to one, which means that one parameter is sufficient to estimate both. The full model has been described in detail by Klauer and Wegener (1998).

Specifically, a possible way to understand the MPT model is that upon perception of a statement in the test phase, participants first try to remember whether they have seen the statement in the learning phase (item discrimination). If they do not remember the statement, they will respond "no" to the question whether the statement was shown in the learning phase and the trial is completed. If they do remember the statement, they respond "yes" and next try to remember the speaker of the statement (person discrimination). If they remember the speaker, they will give the correct response. If they do not remember the speaker, their responses depend on their memory of the (un)trustworthiness of the speaker's face (trustworthiness or untrustworthiness encoding). If they remember whether the speaker was trustworthy or untrustworthy, they can at least restrict their guessing to half of the speakers (namely either the trustworthy or untrustworthy speakers), causing systematic guessing errors. MPT modeling estimates the probabilities of the outcomes of these stages (e.g., the probability of remembering the speaker of a statement).[3]

The probabilities were estimated separately for statements made by trustworthy speakers, untrustworthy speakers, and distractor statements (see Klauer & Wegener, 1998). This means that the model would in principle entail three parameters for item discrimination ($D_T$, $D_U$, and $D_N$ where the subscripts *T*, *U*, and *N* stand for trustworthy speakers, untrustworthy speakers, and new statements, respectively). However, a model with all three parameters estimated freely can generally not be identified because it is not sufficiently constrained by the data. Therefore, in line with the MPT model of Klauer and Wegener (1998) we assumed in all analyses that item discrimination parameters were equal ($D_T = D_U = D_N$). A test of this assumption is given in every analysis by assessing the fit of the model with the data. In addition, the model entailed two parameters for person discrimination ($c_T$ and $c_U$) and two parameters for (un)trustworthiness encoding ($d_T$ and $d_U$). In these cases, there were no additional parameters for new statements because person discrimination and trustworthiness encoding can only operate in trials in which old statements were displayed.

We first estimated all parameters together with their confidence intervals. Next, we tested whether the trustworthiness and untrustworthiness encoding parameters together contributed significantly to the model fit by comparing a model where the parameters were estimated freely to a model where the parameters were set to zero. In other words, we tested whether the model would match the data equally well if we assume that no encoding of facial (un)trustworthiness took place. If the model fit was significantly better for the model with freely estimated (un)trustworthiness encoding parameters, we concluded that the parameters contributed significantly to the model fit and thus that trustworthiness or untrustworthiness encoding or both occurred. Only if the parameters jointly contributed significantly to the model fit, the individual trustworthiness and untrustworthiness encoding parameter were tested separately in the same fashion. Notice that probabilities cannot be negative, which means that our test can be significant in only one direction. This means that no a priori hypothesis about the direction of the effect needs to be formulated.

## Results

All participants indicated at the end of the study that they had no difficulty with the English language. Furthermore, manipulation checks showed that trustworthiness ratings were significantly and substantially higher for trustworthy ($M = 5.05$, $SD = 0.73$) compared with untrustworthy faces ($M = 2.77$, $SD = 0.80$), $d = 2.98$, $t(74) = 19.01$, $p < .001$. In addition, participants were significantly more willing to pick houses from trustworthy ($M = 4.64$, $SD = 1.03$) compared with untrustworthy looking brokers ($M = 3.43$, $SD = 0.94$), $d = 1.22$, $t(74) = 6.27$, $p < .001$. Overall, these results confirm that the trustworthiness manipulation was successful and strong.

Next, responses in the 'who said what' task were analyzed using MPT modeling as described above. The MPT model with freely estimated parameters had a satisfactory goodness of fit, $G^2 = 1.01$, $df = 1$, $p = .315$. All parameter estimates and their confidence intervals are given in Table 1. Importantly, the results showed a significant reduction in goodness of fit when constraining the (un)trustworthiness encoding parameters to zero, $\Delta G^2 = 64.97$, $df = 2$, $p < .001$. Likewise, the model fit reduced significantly when constraining only the trustworthiness encoding parameter, $\Delta G^2 = 11.48$, $df = 1$, $p < .001$, and when constraining only the untrustworthiness encoding parameter, $\Delta G^2 = 22.89$, $df = 1$, $p < .001$. Hence, the results showed significant evidence of both trustworthiness and untrustworthiness encoding.

## Discussion

The results indicated that participants encoded facial (un)trustworthiness. These results were obtained even though participants received no impression formation instruction and although their explicit task was merely to read statements. In that sense, (un)trustworthiness encoding was relatively spontaneous. However, those results were obtained under conditions were trustworthiness was relevant to the context (i.e., buying a house from a

[3] The MPT model does not necessarily assume *sequential* processing *stages*. Rather, the nodes in the assumed processing tree reflect *states* of the cognitive system and their *dependencies* upon each other. For example, the MPT model does not necessarily assume that people first try to recall the speaker of a statement and subsequently try to recall the trustworthiness of the speaker if they cannot recall the exact speaker. Instead, the MPT model assumes that trustworthiness will influence responses *if* the speaker is not recalled (thus, a dependency).

Table 1
*Parameter Estimates and 95% Confidence Intervals (CIs) for Study 1*

| Parameter | Estimate | Lower CI | Upper CI |
|---|---|---|---|
| $D_T = D_U = D_N$ | .493 | .473 | .514 |
| $a$ | .505 | .471 | .538 |
| $b$ | .396 | .377 | .415 |
| $c_T$ | .159 | .114 | .204 |
| $c_U$ | .058 | .016 | .100 |
| $d_T$ | .249 | .118 | .380 |
| $d_U$ | .327 | .209 | .444 |

*Note.* The indices indicate whether the speaker of the statement was trustworthy looking (T), untrustworthy looking (U), or whether the statement was new (N).

broker) and made salient (participants rated the trustworthiness of each face prior to the task). In the following study we further investigated the spontaneity of (un)trustworthiness encoding by removing the salience manipulation.

## Study 2

### Method

Study 2 was equivalent to Study 1 with one difference: rather than asking participants to rate the facial trustworthiness of each alleged broker prior to the 'who said what' task (and thus making trustworthiness salient), we asked participants to rate trustworthiness after the task. In other words, Study 2 used a context in which trustworthiness was relevant (buying a house from a broker) but did not additionally make trustworthiness salient (contrary to Study 1). Fifty-one students (35 female) of the Radboud University participated in this study ($M_{age}$ = 22.47, $SD_{age}$ = 3.59). They received five euro or partial course credit as a reward.

### Results

All participants indicated at the end of the study that they had no difficulty with the English language. Furthermore, manipulation checks showed that trustworthiness ratings were significantly and substantially higher for trustworthy ($M$ = 5.29, $SD$ = 0.73) compared with untrustworthy faces ($M$ = 2.96, $SD$ = 0.91), $d$ = 1.64, $t(50)$ = 11.70, $p < .001$. In addition, participants were significantly more willing to pick houses from trustworthy ($M$ = 4.87, $SD$ = 1.00) compared with untrustworthy looking brokers ($M$ = 3.31, $SD$ = 0.94), $d$ = 0.94, $t(50)$ = 6.71, $p < .001$. Overall, these results confirm again that the trustworthiness manipulation was successful, and strong.

Next, responses in the 'who said what' task were analyzed using MPT modeling with freely estimated parameters. The MPT model had a satisfactory goodness of fit, $G^2$ = 0.62, $df$ = 1, $p$ = .429. All parameter estimates and their confidence intervals are given in Table 2. Importantly, the results showed a significant reduction in goodness of fit when constraining the (un)trustworthiness encoding parameters both to zero, $\Delta G^2$ = 74.51, $df$ = 2, $p < .001$. Likewise, the model fit was reduced significantly when constraining only the trustworthiness encoding parameter to zero, $\Delta G^2$ = 11.19, $df$ = 1, $p < .001$, or when constraining only the untrustworthiness encoding parameter to zero, $\Delta G^2$ = 27.27, $df$ = 1, $p < .001$. Hence, the result showed significant evidence of trustworthiness and untrustworthiness encoding.

### Discussion

The results showed evidence of spontaneous encoding of facial (un)trustworthiness cues in a context where trustworthiness is relevant (buying a house from a broker). In fact, the estimates of (un)trustworthiness encoding were relatively similar to those obtained in Study 1 (see Table 1 and 2). This suggests that the salience manipulation in Study 1 had little or no effect, which might reflect that the trustworthiness-relevant context made trustworthiness salient by itself. Alternatively, it is conceivable that people encoded information that is confounded with (un)trustworthiness cues (e.g., attractiveness or masculinity), and that this is why making trustworthiness salient had no strong effect. For these reasons, we next investigated whether spontaneous encoding of facial (un)trustworthiness also occurs in a more neutral context that resembles a situation where a person is encountered in everyday life. In addition, we investigated whether a salience manipulation increases (un)trustworthiness encoding in this context.

## Study 3

When people first encounter another person, they usually start by stating their name and perhaps some general information about themselves. Study 3 mimicked such conditions and investigated to what extent spontaneous trustworthiness encoding occurs. Furthermore, Study 3 had both a condition in which trustworthiness was made salient (*salience* condition) and a condition where trustworthiness was not made salient (*spontaneous* condition). This enabled us to investigate (a) whether people spontaneously encode (un)trustworthiness in a neutral context (*spontaneous* condition), and (b) whether our trustworthiness encoding parameters are influenced by trustworthiness salience (*salience* condition compared with *spontaneous* condition).

### Method

Dutch students ($N$ = 151; 100 female) of the Radboud University participated in this study ($M_{age}$ = 21.62; $SD_{age}$ = 3.30). We created 48 statements that described neutral information about eight imaginary people (e.g., "My flat is next to a supermarket").

Table 2
*Parameter Estimates and 95% Confidence Intervals (CIs) for Study 2*

| Parameter | Estimate | Lower CI | Upper CI |
|---|---|---|---|
| $D_T = D_U = D_N$ | .524 | .499 | .548 |
| $a$ | .489 | .448 | .531 |
| $b$ | .410 | .385 | .435 |
| $c_T$ | .104 | .053 | .154 |
| $c_U$ | .054 | .004 | .105 |
| $d_T$ | .272 | .129 | .416 |
| $d_U$ | .428 | .294 | .562 |

*Note.* The indices indicate whether the speaker of the statement was trustworthy looking (T), untrustworthy looking (U), or whether the statement was new (N).

In addition, 46 distractor statements were created that also gave information about imaginary people. Each statement included general and relatively neutral information (e.g., name, age, city of residence, use of public transport, etc.; see material on Open Science Framework).

In addition, participants were assigned at random to one of two between-subjects conditions. In the *salience* condition ($n = 77$), participants were asked to judge the trustworthiness of each speaker's face *prior* to the 'Who said what' task. In the *spontaneous* condition ($n = 74$), participants were asked to judge the trustworthiness *after* the 'Who said what' task. The purpose of asking for trustworthiness judgments prior to the 'Who said what' task was to draw attention to the trustworthiness of the speakers and thereby to influence the (un)trustworthiness encoding parameters. No further questions were asked (aside from demographical questions). Everything else was identical to Studies 1 and 2.

## Results

First, we rechecked participants' trustworthiness ratings of the speakers collapsed over the salience and spontaneous condition. Overall, trustworthy faces ($M = 5.31$, $SD = 0.72$) were judged as substantially more trustworthy looking than untrustworthy faces ($M = 2.81$, $SD = 0.89$), $d = 2.11$, $t(150) = 25.93$, $p < .001$. The same result was obtained within only the salience condition, $d = 2.43$, $t(76) = 21.35$, $p < .001$, and within only the spontaneous condition, $d = 1.86$, $t(73) = 15.97$, $p < .001$. Hence, the trustworthiness manipulation of the speaker's faces appeared to be successful and strong in all conditions.

Next, we fitted an MPT model on the whole data from the 'who said what' task with separate multinomial processing trees for the two conditions. These trees were structurally equivalent and used the same parameters with the exception that there were separate (un)trustworthiness encoding parameters for the salience and the spontaneous condition. The model had a satisfactory goodness of fit, $G^2 = 5.77$, $df = 7$, $p = .567$. All parameter estimates are given in Table 3.

Did participants spontaneously encode facial (un)trustworthiness? To answer this question, we constrained the (un)trustworthiness encoding parameters to zero in the spontaneous condition. This caused a significant reduction in the model fit, $\Delta G^2 = 338.10$, $df = 2$, $p < .001$. Likewise, the model fit was significantly reduced when constraining only the trustworthiness encoding parameter to zero, $\Delta G^2 = 50.03$, $df = 1$, $p < .001$, or when constraining only the untrustworthiness encoding parameter to zero, $\Delta G^2 = 34.84$, $df = 1$, $p < .001$. Hence, we observed significant evidence of both trustworthiness and untrustworthiness encoding in the spontaneous condition.

Table 3
*Parameter Estimates and 95% Confidence Intervals (CIs) for Study 3 With Separate (Un)Trustworthiness Encoding Parameters for the Salience and the Spontaneous Condition*

| Parameter | Estimate | Lower CI | Upper CI |
|---|---|---|---|
| $D_T = D_U = D_N$ | .674 | .663 | .686 |
| $a$ | .449 | .390 | .509 |
| $b$ | .113 | .101 | .125 |
| $c_T$ | .303 | .278 | .328 |
| $c_U$ | .174 | .150 | .199 |
| $d_T$ *(salience)* | .477 | .388 | .567 |
| $d_T$ *(spontaneous)* | .383 | .285 | .482 |
| $d_U$ *(salience)* | .454 | .350 | .558 |
| $d_U$ *(spontaneous)* | .351 | .233 | .469 |

*Note.* The indices indicate whether the speaker of the statement was trustworthy looking (T), untrustworthy looking (U), or whether the statement was new (N).

Did the salience manipulation increase (un)trustworthiness encoding? To answer this question, we constrained the (un)trustworthiness encoding parameters to be equal across conditions. This caused a significant reduction in the model fit, $\Delta G^2 = 8.98$, $df = 2$, $p = .011$. The same was true if only trustworthiness encoding was constrained to be equal across conditions, $\Delta G^2 = 4.36$, $df = 1$, $p = .037$, and if only untrustworthiness encoding was constrained to be equal across conditions, $\Delta G^2 = 4.61$, $df = 1$, $p = .032$. These results indicate that both trustworthiness and untrustworthiness encoding were not equal in these conditions. More specifically, both trustworthiness and (un)trustworthiness encoding parameter estimates were larger in the salience condition ($d_T = .477$ and $d_U = .454$) compared with the spontaneous condition ($d_T = .383$ and $d_U = .351$). Hence, making trustworthiness salient increased trustworthiness and untrustworthiness encoding.

## Discussion

The results showed evidence for spontaneous encoding of facial (un)trustworthiness cues in a neutral context (i.e., a person introducing him or herself). Moreover, (un)trustworthiness encoding was stronger if (un)trustworthiness was made salient prior to the task compared with a condition where (un)trustworthiness was not made salient. This sensitivity of the (un)trustworthiness encoding parameters to a trustworthiness salience manipulation suggests that these parameters may reflect attention to facial trustworthiness to some degree rather than attention to social information that is confounded with trustworthiness (e.g., attractiveness or masculinity). Taken together, these results further support the conclusion that people spontaneously form trustworthiness impressions based on facial appearance. Studies 1 through 3 showed this using artificial faces with a relatively strong manipulation of facial (un)trustworthiness. A remaining question is whether spontaneous encoding of trustworthiness also occurs based on real faces that differ more subtly in terms of facial trustworthiness. This question was addressed in Study 4.

# Study 4

## Method

Study 4 was equivalent to the spontaneous condition in Study 3 with one difference: instead of using artificial faces, we used real faces. Specifically, we picked the four most trustworthy and four most untrustworthy looking male faces (available on Open Science Framework) from the Radboud Face Database based on supplemented trustworthiness ratings of these faces (Langner et al., 2010). It is important to note that the difference in trustworthiness between these faces is likely to be smaller compared with our artificial faces. Moreover, given that real faces were used, identities could not be counterbalanced in Study 4. The critical question

we aimed to answer was whether (un)trustworthiness encoding is still reliably present with these faces. The study was conducted online (www.prolific.ac), which enabled us to obtain a relatively large and heterogeneous sample of participants. Specifically, 150 Caucasians participated in the study. Two participants were excluded because they indicated that they had problems with understanding the English language or because they did not complete the whole study, leaving 148 participants (57 female; $M_{age} = 31.26$; $SD_{age} = 10.67$).

## Results

Manipulation checks showed that trustworthiness ratings were significantly and substantially higher for trustworthy ($M = 4.98$, $SD = 0.87$) compared with untrustworthy faces ($M = 3.63$, $SD = 1.04$), $d = 1.22$, $t(147) = 14.71$, $p < .001$. This suggests that the preselection of trustworthy and untrustworthy faces was successful. Next, responses in the 'who said what' task were analyzed using MPT modeling with freely estimated parameters. The MPT model had a satisfactory goodness of fit, $G^2 = 3.16$, $df = 1$, $p = .076$. All parameter estimates and their confidence intervals are given in Table 4. Importantly, the results showed a significant reduction in goodness of fit when constraining the (un)trustworthiness encoding parameters both to zero, $\Delta G^2 = 7.54$, $df = 2$, $p < .023$. When testing the trustworthiness and untrustworthiness encoding parameters separately, we found a significant reduction in the model fit when constraining the trustworthiness encoding parameter to zero, $\Delta G^2 = 5.50$, $df = 1$, $p < .019$, but not when constraining the untrustworthiness encoding parameter to zero, $\Delta G^2 = 0.0$, $df = 1$, $p = 1$. Hence, the results showed significant evidence of trustworthiness but not untrustworthiness encoding.

## Discussion

Study 4 investigated the spontaneity of (un)trustworthiness encoding in a neutral context that mimics conditions of a first encounter of a novel person. Importantly, Study 4 employed real faces that differed less strongly in terms of facial trustworthiness (difference in trustworthiness ratings: $d = 1.22$) compared with the faces employed in Studies 1–3 (difference in trustworthiness ratings: $d = 2.98$, $d = 1.64$, and $d = 2.11$, respectively). The results showed significant evidence of trustworthiness but no evidence of untrustworthiness encoding.

Table 4
*Parameter Estimates and 95% Confidence Intervals (CIs) for Study 4*

| Parameter | Estimate | Lower CI | Upper CI |
|---|---|---|---|
| $D_T = D_U = D_N$ | .526 | .513 | .539 |
| $a$ | .505 | .471 | .539 |
| $b$ | .237 | .224 | .250 |
| $c_T$ | .296 | .267 | .325 |
| $c_U$ | .464 | .435 | .494 |
| $d_T$ | .124 | .019 | .230 |
| $d_U$ | .000 | −.122 | .122 |

*Note.* The indices indicate whether the speaker of the statement was trustworthy looking (T), untrustworthy looking (U), or whether the statement was new (N).

The former supports the conclusion that participants *spontaneously* encoded that a perceived face appears trustworthy. In contrast, the interpretation of untrustworthiness encoding parameter is less straightforward. What is remarkable is that the untrustworthiness encoding parameter was not merely estimated to be small but literally zero. One possible explanation for this finding is that participants did not encode facial untrustworthiness. However, an alternative explanation is that facial untrustworthiness facilitated person discrimination, and that the untrustworthiness encoding parameter may therefore underestimate the true extend of untrustworthiness encoding. This is because (un)trustworthiness encoding is only estimated in trials where person discrimination failed (see Figure 2). Consequently, every trial in which detecting facial untrustworthiness caused accurate person discrimination is not taken into account in the estimation of the untrustworthiness encoding parameter. As a result, encoding of facial untrustworthiness could potentially have become indiscernible by a facilitative effect on person discrimination. This interpretation converges with the exploratory observation (see Table 4) that person discrimination was larger for untrustworthy faces ($c_u = .464$) compared with trustworthy faces ($c_T = .296$; see also Rule, Slepian, & Ambady, 2012).

Taken together, the results support the assumption that people spontaneously encode facial trustworthiness. Moreover, although the results did not show evidence of facial untrustworthiness encoding, the general pattern of the results (i.e., when taking person discrimination into account) suggests that this could be due to limitations of the MPT paradigm.

## General Discussion

It is widely assumed among psychologists that people have a strong tendency to spontaneously form trustworthiness impressions from facial appearance. However, existing findings do not fully warrant this assumption, because most existing studies induced an impression formation goal either explicitly (Todorov et al., 2008; Willis & Todorov, 2006) or implicitly (Chang et al., 2010; Rezlescu et al., 2012; Schlicht et al., 2010; Stirrat & Perrett, 2010; van 't Wout & Sanfey, 2008). Moreover, although some studies demonstrated spontaneous neurophysiological responses to facial trustworthiness (Engell et al., 2007; Marzi et al., 2014; Todorov, 2008; Winston, Strange, O'Doherty, & Dolan, 2002), it remains unclear whether this reflects the formation of lasting trustworthiness impressions. Finally, the theoretical plausibility of a spontaneous tendency to infer trustworthiness from facial appearance has been questioned by recent findings. Specifically, it has been found that facial trustworthiness inferences tend to be at chance level accuracy (Rule et al., 2013; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015; but see Slepian & Ames, 2016) and thus do not seem to provide valid (and evolutionary beneficial) information to a perceiver. As such, whether or not people spontaneously form trustworthiness impressions based on facial appearance remained a relatively open question.

To our knowledge, the present studies are the first that tested whether relatively *stable* trustworthiness impressions are formed *spontaneously* from facial appearance. The results of four studies taken together provided evidence for such a tendency. Specifically, the results showed that participant encoded facial (un)trustworthiness if (un)trustworthiness was relevant to the context and made

salient (Study 1), if trustworthiness was relevant to the context without making it salient (Study 2), and if the context mimicked a neutral first encounter of another person (Study 3; spontaneous condition). Furthermore, a saliency manipulation increased (un)trustworthiness encoding in the latter context (Study 3; salience condition). These studies used experimentally controlled artificial faces (Studies 1–3). Finally, we also obtained partial evidence of (un)trustworthiness encoding with more naturalistic varying real faces (Study 4). Taken together, these results provide support for the assumption that people *spontaneously* form relatively *stable* trustworthiness impression from facial appearance. As such, these results contribute to closing an important gap in the empirical social perception literature.

## Societal Implications

Previous studies have shown that facial trustworthiness influences important behavioral outcomes in contexts that require making trustworthiness related decisions (Chang et al., 2010; Porter et al., 2010; Rezlescu et al., 2012; Schlicht et al., 2010; Stirrat & Perrett, 2010; Todorov et al., 2015; van 't Wout & Sanfey, 2008). Our results suggest that facial trustworthiness may also be *encoded* in relatively neutral contexts in which trustworthiness is not explicitly relevant. This further suggests that *behavioral outcomes* of facial trustworthiness on behavior may be relatively independent of the context in which a face is perceived. For example, even if a person is initially encountered in a neutral context (e.g., in a supermarket) and only later a decision needs to be made about the person (e.g., whether to invite the person for an interview based on a CV without a picture), facial trustworthiness may influence the decision. This further emphasizes that facial trustworthiness may have pervasive consequences in everyday life.

## Methodological Implications

Our studies also demonstrate the broad applicability of the 'who said what' paradigm. Originally, the 'who said what' paradigm was conceived of as a method for detecting spontaneous categorization into discrete classes (e.g., male and female; Taylor et al., 1978). In contrast, trustworthiness and untrustworthiness do not necessarily constitute discrete classes but could in principle be seen as endpoints of the same social dimension (i.e., trustworthiness). For this reason, it was not entirely clear a priori whether the 'who said what' paradigm can be used to measure trustworthiness encoding. Our findings show that the 'who said what' paradigm is sensitive to (un)trustworthiness encoding. This converges with various other studies in which the "Who said what" paradigm was applied to various different cues (Klauer & Wegener, 1998). Taken together, this suggest that the 'who said what' paradigm may be conceived of as a method to measure (spontaneous) cue encoding in general, and may thus be more widely applicable than originally assumed.

## Limitations

Facial trustworthiness cues are intrinsically confounded with other facial cues such as attractiveness, age, and sex (Todorov et al., 2008). As such, it is conceivable that our results reflect encoding of other information than (un)trustworthiness. This is an inevitable limitation that is shared by previous studies (Chang et al., 2010; Engell et al., 2007; Rezlescu et al., 2012; Schlicht et al., 2010; Stirrat & Perrett, 2010; Todorov, 2008; van 't Wout & Sanfey, 2008; Willis & Todorov, 2006; Winston et al., 2002). We attempted to minimize this limitation by creating artificial faces that are manipulated in terms of (un)trustworthiness while keeping variations on other dimensions as constant as possible. In addition, the results showed that the obtained effect gets stronger if (un)trustworthiness is made salient but only if it is not already salient due to a (un)trustworthiness-relevant context. Although this does not fully rule out alternative explanations (e.g., encoding of age cues), the pattern of these results suggests that the obtained effects reflect encoding of facial (un)trustworthiness to some degree.

Another limitation is that we relied exclusively on the 'who said what' paradigm. This paradigm has the strength that it does not explicitly induce an impression formation goal, and does not require mentioning (un)trustworthiness to participants. Furthermore, this paradigm has the strength that it measures whether social cues are not only detected but also encoded in memory. Nevertheless, a limitation is that this paradigm assumes that the underlying processes are uncorrelated (Klauer & Wegener, 1998). In particular, if there is a correlation between person discrimination and (un)trustworthiness encoding, the amount of (un)trustworthiness encoding is imperfectly estimated. This is because (un)trustworthiness encoding is estimated exclusively based on trials in which person discrimination failed (see Figure 2) and does not take the amount of (un)trustworthiness encoding into account that happened in trials where person discrimination succeeded. This is particularly important for the interpretation of the results of Study 4 where person discrimination was relatively high. It is conceivable that the reason we found evidence for trustworthiness encoding but no evidence for untrustworthiness encoding in Study 4 is that people tend to remember untrustworthy faces (i.e., successful person discrimination). To the extent that this is the case, the results of Study 4 underestimate the amount of untrustworthiness encoding. Importantly, if anything this possibility strengthens the conclusion that people may spontaneously encode facial (un-)trustworthiness.

## Future Research

Future research may complement our work by investigating the spontaneous encoding of other facial cues (e.g., dominance). Furthermore, another possible direction is to investigate how facial cues interact with behavioral cues. Previous studies showed that people form initial trustworthiness impressions based on facial appearance in a trust game but gradually update this impression based on incoming behavioral information (Chang et al., 2010). An open question is how facial appearance interacts with behavioral cues in contexts where trustworthiness is not salient. For example, it is conceivable that updating an initial face-based trustworthiness impression happens mainly if people have the goal to form an accurate trustworthiness impression but not when trustworthiness impressions are formed incidentally.

Another open question is to what extent people spontaneously encode these trustworthiness cues if they observe dynamically moving faces. A main explanation for the tendency to infer trustworthiness from the structure a face is that people may confuse

facial expressions (which may provide valid cues to trustworthiness) with facial structure (which may not provide any valid cues to trustworthiness; Todorov, 2008). For example, some people may have a facial structure that makes it appear as if these people are smiling (a trustworthiness cue), whereas other people may have a facial structure that makes it appears as if these people are frowning (an untrustworthiness cue). However, social perceivers may be able to disentangle facial expressions and facial structure more effectively when observing dynamically moving faces. As a result, they may be less inclined to encode (alleged) trustworthiness cues in facial structure in this situation. Future research may investigate this by employing videos of moving faces while independently varying facial structure and dynamic facial expressions.

## Conclusion

It is widely assumed among psychologists that people spontaneously "judge a book by its cover": they infer how trustworthy a perceived person is based on the person's facial appearance. However, the existing findings did not fully warrant this assumption. Our results provide empirical support for the assumption that people spontaneously infer trustworthiness from facial appearance, and thus contribute to closing this important gap in the literature. In particular, our results suggest that facial (un)trustworthiness is spontaneously encoded in memory. This further emphasizes the pervasive consequences facial trustworthiness may have in our daily life.

## References


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,* 87–105. http://dx.doi.org/10.1016/j.cogpsych.2010.03.001

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. http://dx.doi.org/10.1162/jocn.2007.19.9.1508

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the "who said what?" paradigm. *Journal of Personality and Social Psychology, 75,* 1155–1178. http://dx.doi.org/10.1037/0022-3514.75.5.1155

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion, 24,* 1377–1388. http://dx.doi.org/10.1080/02699930903485076

Maddox, K. B., & Gray, S. (2002). Cognitive representations of Black Americans: Reexploring the role of skin tone. *Personality and Social Psychology Bulletin, 28,* 250–259. http://dx.doi.org/10.1177/0146167202282010

Marzi, T., Righi, S., Ottonello, S., Cincotta, M., & Viggiano, M. P. (2014). Trust at first sight: Evidence from ERPs. *Social Cognitive and Affective Neuroscience, 9,* 63–72. http://dx.doi.org/10.1093/scan/nss102

Mende-Siedlecki, P., Said, C. P., & Todorov, A. (2013). The social evaluation of faces: A meta-analysis of functional neuroimaging studies. *Social Cognitive and Affective Neuroscience, 8,* 285–299. http://dx.doi.org/10.1093/scan/nsr090

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 11087–11092. http://dx.doi.org/10.1073/pnas.0805664105

Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law, 16,* 477–491. http://dx.doi.org/10.1080/10683160902926141

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,* e34293. http://dx.doi.org/10.1371/journal.pone.0034293

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95,* 318–339. http://dx.doi.org/10.1037/0033-295X.95.3.318

Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology, 104,* 409–426. http://dx.doi.org/10.1037/a0031050

Rule, N. O., Slepian, M. L., & Ambady, N. (2012). A memory advantage for untrustworthy faces. *Cognition, 125,* 207–218. http://dx.doi.org/10.1016/j.cognition.2012.06.017

Said, C. P., Dotsch, R., & Todorov, A. (2010). The amygdala and FFA track both social and non-social face dimensions. *Neuropsychologia, 48,* 3596–3605. http://dx.doi.org/10.1016/j.neuropsychologia.2010.08.009

Schlicht, E. J., Shimojo, S., Camerer, C. F., Battaglia, P., & Nakayama, K. (2010). Human wagering behavior depends on opponents' faces. *PLoS ONE, 5,* e11663. http://dx.doi.org/10.1371/journal.pone.0011663

Singmann, H., & Kellen, D. (2013). MPTinR: Analysis of multinomial processing tree models in R. *Behavior Research Methods, 45,* 560–575. http://dx.doi.org/10.3758/s13428-012-0259-0

Slepian, M. L., & Ames, D. R. (2016). Internalized impressions: The link between apparent facial trustworthiness and deceptive behavior is mediated by targets' expectations of how they will be judged. *Psychological Science, 27,* 282–288. http://dx.doi.org/10.1177/0956797615594897

Sofer, C., Dotsch, R., Wigboldus, D. H. J., & Todorov, A. (2015). What is typical is good: The influence of face typicality on perceived trustworthiness. *Psychological Science, 26,* 39–47.

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science, 21,* 349–354. http://dx.doi.org/10.1177/0956797610362647

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology, 36,* 778–793. http://dx.doi.org/10.1037/0022-3514.36.7.778

Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences, 1124,* 208–224. http://dx.doi.org/10.1196/annals.1440.012

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion, 13,* 724–738. http://dx.doi.org/10.1037/a0032335

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. http://dx.doi.org/10.1146/annurev-psych-113011-143831

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. http://dx.doi.org/10.1016/j.tics.2008.10.001

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. http://dx.doi.org/10.1037/0022-3514.83.5.1051

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology, 87,* 482–493. http://dx.doi.org/10.1037/0022-3514.87.4.482

Uleman, J. S., Hon, A., Roman, R. J., & Mokowitz, G. B. (1996). On-line evidence for spontaneous trait inferences at encoding. *Personality and*

*Social Psychology Bulletin, 22,* 377–394 http://dx.doi.org/10.1177/0146167296224005

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. http://dx.doi.org/10.1016/S0065-2601(08)60239-7

Van Overwalle, F., Drenth, T., & Marsman, G. (1999). Spontaneous trait inferences: Are they linked to the actor or to the action? *Personality and Social Psychology Bulletin, 25,* 450–462. http://dx.doi.org/10.1177/0146167299025004005

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. http://dx.doi.org/10.1016/j.cognition.2008.07.002

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17,* 592–598. http://dx.doi.org/10.1111/j.1467-9280.2006.01750.x

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience, 5,* 277–283. http://dx.doi.org/10.1038/nn816

Zebrowitz, L. A., Fellous, J.-M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. *Personality and Social Psychology Review, 7,* 194–215. http://dx.doi.org/10.1207/S15327957PSPR0703_01

Received October 1, 2015
Revision received July 21, 2016
Accepted July 21, 2016 ■