

Is the Face a Window to the Soul? Investigation of the Accuracy of Intuitive Judgments of the Trustworthiness of Human Faces

Stephen Porter, Laura England, Marcus Juodis, Leanne ten Brinke, and Kevin Wilson
Dalhousie University

Although trustworthiness judgments based on a stranger's face occur rapidly (Willis & Todorov, 2006), their accuracy is unknown. We examined the accuracy of trustworthiness judgments of the faces of 2 groups differing in trustworthiness (Nobel Peace Prize recipients/humanitarians vs. *America's Most Wanted* criminals). Participants viewed 34 faces each for 100 ms or 30 s and rated their trustworthiness. Subsequently, participants were informed about the nature of the 2 groups and estimated group membership for each face. Judgments formed with extremely brief exposure were similar in accuracy and confidence to those formed after a long exposure. However, initial judgments of untrustworthy (criminals') faces were less accurate ($M = 48.8\%$) than were those of trustworthy faces ($M = 62.7\%$). Judgment accuracy was above chance for trustworthy targets only at Time 1 and slightly above chance for both target types at Time 2. Participants relied on perceived kindness and aggressiveness to inform their rapidly formed intuitive decisions. Thus, intuition plays a minor facilitative role in reading faces.

Keywords: face, trustworthiness, intuition, credibility

The face acts as the display board on which emotions and intentions are communicated, and is scrutinized by others during social interactions and encounters with others (Martelli, Majib, & Pelli, 2005). A stranger's facial characteristics play a powerful role in rapid observer inferences about his/her global and specific trait features (e.g., Ambady, Krabbenhoft, & Hogan, 2006; Todorov, Mandisodza, Goren, & Hall, 2005), including personality features such as agreeableness, extraversion, and emotional stability (Penton-Voak, Pound, Little, & Perrett, 2006). Further, there is some evidence that these assessments can be correct. Personality traits such as extraversion and conscientiousness can be judged at an accuracy level greater than chance simply by viewing the target's face (Little & Perrett, 2007). Thus, evidence suggests that complex information about one's behavioural tendencies can be communicated via the face and, to a limited extent, accurately assessed by observers.

Among the most important inferences that must be made on encountering an unknown person is whether he or she is trustworthy or potentially dangerous. This basic discrimination of friend from foe may have been one of the earliest interpersonal judgments to evolve (e.g., Cosmides & Tooby, 1992). To inform one's best course of action for survival, such decisions had to occur rapidly and were based largely on facial appearance and expression (e.g., Pessoa, Japee, Sturman, & Ungerleider, 2005). Willis and Todorov (2006) manipulated participants' exposure time to strangers' faces to investigate the minimal conditions under which face-based judgments are formulated. They found that a fleeting (100-ms) exposure to faces allowed for the formation of

enduring judgments of trustworthiness; extending time exposure only served to increase confidence in the rapidly formed judgments.

But how accurate are intuitive, face-based impressions of trustworthiness? The answer to this question has enormous applied implications. For example, the face plays a major role in legal decision making. Individuals charged with crimes considered to be congruent with their facial appearance are more likely to be found guilty in mock jury studies (Dumas & Teste, 2006). Field research has highlighted the effect of facial features on judicial outcomes, suggesting that "babyfacedness," symmetry, and attractiveness are associated with perceptions of honesty (Bull, 2006; Bull & Vine, 2003; Zebrowitz, Voinescu, & Collins, 1996). Laypeople and police officers alike hold stereotypes regarding the types of crimes likely to be committed by an individual with particular facial characteristics (e.g., Bull & Green, 1980; Macrae & Shepherd, 1989). Inferences based on the face are a powerful force in the courtroom. In *R. v. B. (K.G.)* (1993), the Supreme Court of Canada concluded that judges and jurors must be able to view a witness's face because it is an important indicator of credibility not available from a written transcript. In the airport context, the facial appearance of passengers is heavily relied on to estimate the likelihood of dangerousness. The 9/11 attacks and other terrorist plots prompted the United States' transportation agency to develop training aimed at helping staff identify threats by "reading" faces (see Porter & ten Brinke, 2008). It plans to train hundreds of officers in this approach and deploy them at all major airports by 2008 (Lipton, 2006).

Clearly, interpersonal assessment based on others' facial characteristics occurs frequently, can hold major consequences for decision making, and is a skill that, subjectively, appears to be accurate. Humans pay attention to strangers' faces because they think they can decide what they are like on the basis of this impression. One possibility is that this evaluation process evolved

Stephen Porter, Laura England, Marcus Juodis, Leanne ten Brinke, and Kevin Wilson, Department of Psychology, Dalhousie University.

Correspondence concerning this article should be addressed to Stephen Porter, Department of Psychology, Dalhousie University, Halifax, NS, Canada B3H 4J1. E-mail: sbporter@dal.ca

effectively in our distant past, maximising the chances of survival. That is, perhaps these automatic judgments are generally accurate. Indeed, reading a face can sometimes be a simple task. Confronted with a threatening face displaying anger signs, such as lowered brows, flared nostrils, and “flashing eyes” (Darwin, 1872), one might wisely make one’s escape.

Beyond situations involving immediate threats, it is clear that humans are remarkably perceptive in reading certain target characteristics from faces, in addition to those already mentioned. For example, after viewing photographs of male faces, women can accurately identify those who have given higher preference ratings to pictures of infants (Roney, Hanson, Durante, & Maestripieri, 2006). However, it is possible that trustworthiness judgments, specifically, may be highly flawed. In the absence of objective information concerning past behaviour or character that may be relevant to one’s trustworthiness, judgments of a stranger’s face are inherently intuitive. Despite the value attached to such intuitive judgments, there is evidence that the use of intuition is not a valid strategy for evaluating credibility or trustworthiness more generally. For example, Porter, Woodworth, and Birt (2000) found that Canadian parole officers performed worse than chance in a task requiring them to judge whether videotaped speakers were lying. Self-reported reliance on intuition (or gut instinct) and accuracy in detecting deception were inversely related. Many participants reported that there was something about the person’s face that had contributed to their (faulty) judgments.

Another reason to hypothesise that this process is fallible is that there are individual differences in observer inferences about the trustworthiness of a common face (Adolphs, 2002; Richell et al., 2005). Further, facial appearance and expression can be manipulated to mask or conceal one’s true intentions or character. Indeed, many serious criminals are proficient deceivers and highly adept at gaining the trust of others (e.g., Porter & Woodworth, 2007). In particular, psychopathic individuals are successful in manipulating others by their charm and “honest face” (e.g., Hare, 2003). Thus, trustworthiness judgments may represent mere guesses, and faces of highly untrustworthy individuals (such as dangerous criminals) may be particularly difficult to evaluate correctly. Thus, although Willis and Todorov (2006) appeared to assume that rapid trustworthiness judgments evolved to be accurate, this assumption is unfounded because the dispositional features of the targets in their study were unknown. Further, Bar, Neta, and Linz (2006) found that individuals evaluate the threat posed by another on the basis of his or her facial characteristics and that these assessments are reliably formed after only 39 ms. Like Willis and Todorov (2006), they suggested that the speed of this process has evolved to enhance survival; however, the actual threat posed by their target faces again was unknown.

The present study is the first to examine the accuracy of initial impressions of trustworthiness from the faces of known targets presumably differing on the characteristic. Specifically, we compared judgments of trustworthiness assigned to a group of dangerous, untrustworthy individuals (*America’s Most Wanted* criminals) and a group of relatively virtuous, trustworthy individuals (Nobel Peace Prize winners and Order of Canada humanitarians). If intuitive judgments of trustworthiness are valid, the discrimination of the most and least trustworthy individuals in society should be possible. Although it is likely that everyone’s honesty fluctuates (because lying is such an important part of social interaction; Vrij,

2008), we were interested in examining judgments of the broader trait of trustworthiness, including honesty and dangerousness. Thus, defining groups on the basis of being a violent fugitive or an acknowledged advocate of peace and philanthropy is an appropriate approach to address the present research question.

Our methodological approach was to present target faces varying in trustworthiness (opposite ends of the spectrum) individually for either a very brief or an extended exposure time. For each target, participants were first asked whether the target was generally trustworthy or untrustworthy and to provide ratings on trustworthiness, kindness, and aggressiveness and corresponding confidence ratings for each. Subsequently, participants were informed about the general nature of the two groups and asked to reevaluate each face, this time deciding on the target’s group membership.

Method

Participants

Undergraduate students ranging in age from 18 to 35 years ($M = 21.67$, $SD = 2.73$) participated in the pilot and main studies in return for course credit. For the pilot study, we recruited 20 participants (7 men and 13 women) to obtain prestudy ratings for each target face. We recruited 30 participants (6 men and 24 women) to participate in the main study; equal numbers made trait inferences after viewing photographs of the targets for either 100 ms or 30 s.

Face Stimuli and Apparatus

The face stimuli (targets) were 34 black-and-white photographs of adult male faces. Seventeen of the targets were relatively high in trustworthiness. Specifically, they had either received the Nobel Peace Prize or the Order of Canada and had been acknowledged as paragons in their devotion to humanity, peace, and society. The other 17 targets were criminals depicted on the Web site of America’s Most Wanted, an organisation that profiles fugitives eluding justice for extremely serious crimes, including those currently on the FBI’s Ten Most Wanted Fugitives list. Although it is not possible to know the absolute trustworthiness of these targets (or of anyone, for that matter), we assume that individuals in the two groups will generally differ (and we expect will differ greatly) in their relative level of trustworthiness. Permission to copy and use the photographs was granted by the Web site operators and agencies who own the images. Each photograph was cropped to expose only the face (no hair) and was mounted on a grey background, following the method of Bar et al. (2006). The photographs were presented on a 17-in. (43.2-cm) computer screen. The targets were matched on facial hair, facial expressions, and ethnicity. This procedure ensured that judgments were based specifically on the face rather than on obvious external cues (e.g., prison jumpsuits worn by America’s Most Wanted criminals vs. suits worn by most of the high trustworthy targets). Participants in the pilot study viewed all target faces (randomly presented) for an unlimited amount of time and rated each of them on various characteristics (see below). To obtain the trait ratings of interest, we administered four self-report questionnaires: (a) Prestudy Stimulus Rating Scale, (b) Assessment of Traits Inventory—Section One (ATI-1), (c) Assessment of Traits Inventory—Section Two

(ATI-2), and (c) Trait Emotional Intelligence Questionnaire Short Form (TEIQue-SF; Petrides & Furnham, 2006).

Prestudy Stimulus Rating Scale. This scale was designed to obtain pilot ratings for the target faces on the following characteristics: attractiveness, babyfacedness, symmetry, kindness, aggression, age, ethnicity, and familiarity. It consisted of trait-judgment questions answered using either a scale ranging from 1 (*not at all*) to 7 (*extremely*; e.g., "Please rate this person's attractiveness on the following scale") or a written estimate (e.g., "Please estimate the age of this person in years"). In addition, each rating was accompanied by a confidence rating ranging from 1 (*not at all confident*) to 7 (*extremely confident*).

ATI-1. The ATI-1, a rating sheet designed for this study, consisted of trait judgments made after exposure to a target face: trustworthiness, aggression, and kindness. Responses ranged from 1 (*not at all*) to 7 (*extremely*), followed by a confidence rating. Furthermore, one dichotomous trait judgment question was asked: "Is this person generally trustworthy?" (yes or no).

ATI-2. The ATI-2, also designed for this study, included three questions to assess final trait judgments after participants were informed about the nature and base rate of the two groups. Participants were asked to indicate their estimate of group membership and confidence (rated on a scale ranging from 1 to 7).

TEIQue-SF. The TEIQue-SF is a 30-item inventory that assesses the global trait of emotional intelligence, including the facets of well-being, self-control, emotionality, and sociability. Each item is rated on a 7-point Likert scale. The TEIQue-SF has good reliability, validity, and internal consistency (Petrides & Furnham, 2003).

Procedure

For the main study, individual participants were randomly assigned to one of the two time exposure trials, 100 ms or 30 s, and one of two randomly created presentation orders. Participants were informed that the study focused on their first impressions of photographed stranger faces. In the unlikely event that they recognized a face, they were asked to alert the experimenter. All participants had up to 30 s following the target presentation to complete the ATI-1 (i.e., scoring form for ratings of trustworthiness, kindness, and aggressiveness), after which the next target was presented. The TEIQue-SF was then completed before the second target presentation. Following completion of the TEIQue-SF, participants were informed of the existence of the two categories of faces (i.e., fugitives and award winners) and asked to reevaluate each target, attempting to correctly categorize each face. During this final stimulus presentation, all targets were shown for 10 s, after which participants were given 30 s to complete additional ratings on the ATI-2.

Results

Pilot Ratings of the Trustworthy and Untrustworthy Targets

We obtained pilot ratings to examine whether the two target groups differed on a variety of characteristics. We conducted a multivariate analysis of variance to compare trustworthy and untrustworthy faces on ratings of babyfacedness, facial symmetry,

attractiveness, perceived kindness, perceived aggression, and estimated age. There were no differences on these pilot ratings, $F(6, 14) = 2.62, p > .05$. Further, none of the 34 targets were recognized by any participant.

Can We Accurately Infer Trustworthiness From the Human Face?

Within the main study, a series of one-sample *t* tests (a priori Bonferroni correction places the significance level at $p < .006$) revealed that in the brief exposure condition, only untrustworthy faces were accurately assessed at rates better than chance during the final viewing, although the difference was only marginally significant when considering the adjusted criterion, $t(14) = 3.06, p = .008$. In terms of accuracy, untrustworthy faces were assessed at approximately 59.61% accuracy. Within the brief exposure group, accuracy was never greater than chance, save for the above result.

For those in the extended exposure group, accuracy for trustworthy faces at both initial, $t(14) = 4.72, p < .0005$ (63.14%), and final judgment times, $t(14) = 5.15, p < .0005$ (64.31%), was significantly greater than chance. Accuracy for untrustworthy faces was never significantly greater than chance across any of the extended exposure conditions (all $ps > .05$).

Consideration of Covariates

Despite the lack of significant differences in secondary traits across groups in the pilot study, two traits (i.e., kindness and aggressiveness) were retained in the main study because of their theoretical importance to intuitive judgments. One-way analyses of variance using target type as the independent variable revealed that trustworthy faces ($M = 4.48, SD = 0.44$) were rated as being significantly more kind than untrustworthy faces ($M = 4.08, SD = 0.36$), $F(1, 29) = 38.65, p < .001, \eta_p^2 = .57$. Trustworthy faces ($M = 3.85, SD = 0.58$) were also rated as being significantly less aggressive than untrustworthy faces ($M = 3.51, SD = 0.59$), $F(1, 29) = 13.01, p < .01, \eta_p^2 = .31$. These differences suggest that the extent to which participants discriminate trustworthy and untrustworthy faces on the basis of kindness and aggressiveness should be a controlled variable in further analyses.

Validity of the Trustworthiness Judgments

We performed a 2 (exposure time: brief vs. extended) \times 2 (judgment time: initial vs. final) \times 2 (target type: trustworthy vs. untrustworthy) mixed analysis of covariance (ANCOVA) with the mean accuracy scores (number of accurate assessments for trustworthy and untrustworthy targets) as the dependent variable. The differences between each participant's ratings of kindness and aggressiveness for trustworthy and untrustworthy faces (i.e., the extent to which these variables were used to discriminate faces) were entered as covariates. The mixed ANCOVA revealed a main effect of target type, $F(1, 26) = 4.70, p < .05, \eta_p^2 = .15$, such that trustworthy faces ($M = 10.38, SD = 1.80$, percentage correct = 61.06%) were judged correctly significantly more often than untrustworthy faces ($M = 8.95, SD = 1.77$, percentage correct = 52.65%). There was also a main effect of judgment time, $F(1, 26) = 9.49, p < .01, \eta_p^2 = .27$, wherein initial judgments ($M =$

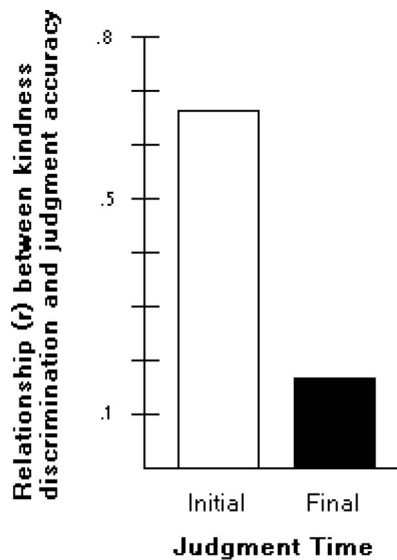


Figure 1. The relationship (expressed in correlations) between kindness discrimination and judgment accuracy across judgment times.

9.48, $SD = 1.00$, percentage correct = 55.77%) were significantly less accurate than final judgments ($M = 9.85$, $SD = 1.41$, percentage correct = 57.94%). There was also a significant judgment time \times kindness discrimination interaction, $F(1, 26) = 8.37$, $p < .01$, $\eta_p^2 = .24$. There were no other main effects or significant interactions (all $ps > .05$).

To explore the interaction between kindness discrimination and judgment time, we ran correlations between each participant's level of kindness discrimination and accuracy at both judgment times. These revealed a significant correlation between kindness discrimination and accuracy during initial judgments, $r(29) = .66$, $p < .001$, but not final judgments, $r(29) = .17$, $p = .38$ (see Figure 1).

Signal Detection Analysis

We converted accuracy scores into D' (sensitivity to untrustworthiness) and C (decision bias) and conducted similar analyses to those above. A 2 (exposure time) \times 2 (judgment time) multivariate analysis of covariance was conducted with D' and C as the dependent variables and kindness and aggressiveness discrimination as covariates. These analyses revealed a multivariate main effect of judgment time, $F(2, 25) = 5.85$, $p < .01$, $\eta_p^2 = .32$, and a significant multivariate interaction between judgment time and kindness discrimination, $F(2, 25) = 4.46$, $p < .05$, $\eta_p^2 = .26$. At the univariate level, the main effect of judgment time was significant for D' , $F(1, 26) = 9.00$, $p < .01$, $\eta_p^2 = .26$, such that the ability to discriminate between trustworthy and untrustworthy faces was lower at initial judgments ($M = .24$, $SD = .24$) than at final judgments ($M = .30$, $SD = .31$). The univariate interaction between judgment time and kindness discrimination also was significant for D' , $F(1, 26) = 9.19$, $p < .01$, $\eta_p^2 = .26$. To explore this interaction, we conducted correlations across initial and final judgment times. These analyses revealed that kindness discrimination was significantly correlated with the ability to discriminate trustworthy and untrustworthy faces during initial judgments,

$r(29) = .66$, $p < .001$, but not final judgments, $r(29) = .15$, $p = .42$. No other multivariate effects or interactions were significant.

Because the literature indicates a tendency for participants to demonstrate a trust bias (i.e., they tend to rate both truth tellers and liars as being honest; O'Sullivan, Ekman, & Friesen, 1988), we conducted analyses to compare C with its zero bias statistic (0). Values may range from -1 to 1 , with values less than zero indicating a tendency to respond liberally and those above zero indicating a conservative bias (claiming faces are untrustworthy or trustworthy, respectively). We conducted one-sample t tests, which revealed that the average C was greater than zero during initial judgments ($M = .21$, $SD = .41$), $t(29) = 2.74$, $p = .01$, but not final judgments ($M = .04$, $SD = .22$), $t(29) = 1.01$, $p = .29$, indicating that participants were conservative in declaring faces untrustworthy during their initial judgments, but were unbiased when making their final judgments.

Confidence in the Trustworthiness Judgments

To examine whether confidence was influenced by exposure time, judgment time, or target type, we conducted a 2 \times 2 \times 2 mixed ANCOVA with the mean confidence scores for trustworthy and untrustworthy judgments as the dependent variable and both kindness and aggression discrimination entered as covariates. The mixed ANCOVA revealed no main effects; however, the judgment time \times kindness interaction was significant, $F(1, 26) = 5.61$, $p < .05$, $\eta_p^2 = .17$. To explore the interaction between kindness discrimination and judgment time, we ran correlations between each participant's level of kindness discrimination and judgment confidence at both judgment times. These revealed a significant correlation between kindness discrimination and confidence during initial judgments, $r(29) = .64$, $p < .001$, but not final judgments, $r(29) = .32$, $p = .09$.

Consistency of the Initial and Final Trustworthiness Judgments

Did the face evaluations remain similar following the provision of information concerning the two target categories? The number of consistent judgments (i.e., the number of times a target was consistently judged as trustworthy or untrustworthy at both judgment times) was tabulated for each participant. The number of consistent judgments ranged from 14 to 28 (percentage consistency of 41.18%–82.35%) out of a possible 34, with a mean consistency of 23.03 ($SD = 3.56$), or 67.74%. In other words, on average 32.26% of the judgments changed from Time 1 to Time 2. A 2 (exposure time) \times 2 (target type) mixed ANCOVA (kindness and aggressiveness differentiation entered as covariates) revealed no significant main effects or interactions.

Further Exploration of Intuitive Judgment Strategies

Because kindness discrimination appeared to be a successfully implemented strategy for discerning trustworthiness from target faces, we ran further correlations to explore its usefulness and its relationship with perceived facial aggressiveness (another factor differing between target groups). These analyses revealed a strong negative correlation between kindness and aggressiveness discrimination, $r(29) = -.72$, $p < .001$, indicating that participants who

discriminated between targets on the basis of their perceived kindness did not also discriminate groups on the basis of perceived aggressiveness. Split across judgment times, kindness discrimination was highly correlated with accuracy during initial but not final judgments (see earlier analyses). In contrast, perceived aggressiveness was highly negatively correlated with accuracy during initial judgments, $r(29) = -.54, p < .005$, but not final judgments, although the latter relationship approached significance, $r(29) = -.35, p = .062$.

Did Babyfacedness and Attractiveness Affect Ratings of Trustworthiness?

Bivariate correlations revealed that trustworthiness ratings (made at Time 1 on a scale ranging from 1 to 7) were not significantly related to babyfacedness or attractiveness ratings made by the pilot sample ($ps > .05$). This may have been because of the restricted range of babyfacedness ($M = 3.17, SD = 0.63$) and attractiveness ($M = 2.72, SD = 0.83$) ratings resulting from a young undergraduate population rating faces with a relatively high mean perceived age ($M = 51.51, SD = 4.80$).

Discussion

Do our intuitive judgments accurately evaluate the trustworthiness of others? Is it possible to decide whether a stranger is trustworthy or potentially dangerous simply by attending to his or her facial appearance? On the basis of Willis and Todorov's (2006) findings, we predicted that a brief exposure to a target would lead to the formation of rapid, enduring trustworthiness judgments and that confidence would increase with a longer exposure period. The results supported the first prediction regarding accuracy, but the hypothesis concerning confidence received less support. Trustworthiness judgments formed after 100 ms did not differ in accuracy from those formed after 30 s, providing further support that inferences of trustworthiness are formed instantaneously on seeing a face and endure over time. However, the extended exposure group did not become more confident in their judgments than the brief exposure group. Further, participants were similarly confident in their inferences for the initial and final judgments, suggesting that confidence in trustworthiness ratings may be formed rapidly and endure over time. In Willis and Todorov's (2006) study, each participant viewed the targets at all exposure times (100, 500 and 1,000 ms) and thus were aware of the relative amount of information available during a brief versus longer exposure and may have felt more confident in their judgments formed after viewing the face for a longer period. Here, participants were unable to make relative confidence judgments in brief as compared with long exposure time and, thus, were similarly confident in the ratings.

Of particular interest is the finding that perceived kindness differed between trustworthy and untrustworthy faces and that participants were able to use this cue to successfully discriminate the two groups. However, this strategy was either only effective or only utilised during initial judgments. Kindness discrimination was apparently useful for participants in both the brief and the extended exposure conditions (i.e., there was no Kindness Discrimination \times Exposure Time interaction), suggesting that the strategy is quickly implemented but eventually replaced with more explicitly derived strategies during final judgments. For example,

barring knowledge of the presence of truly untrustworthy targets (i.e., the felons), participants may have relied solely on their own intuitions, including rapid assessments of kindness, to make their decisions. Once aware that some of the targets were highly untrustworthy or dangerous, participants may have adopted more effortful approaches to detecting said targets, ignoring snap judgments of kindness (for instance, by resorting to socially stereotyped methods of catching unsavoury characters; e.g., Hellström & Tekle, 1994).

It is interesting to note that participants were able to use one cue intuitively to discriminate the two groups (i.e., kindness), but entirely ignored another, equally valid cue (i.e., aggressiveness). Indeed, one might imagine that both cues could be used in conjunction with one another to increase assessment accuracy (i.e., use of a "friend or foe?" strategy). However, participants who used kindness to discriminate targets were exceedingly unlikely to also use aggressiveness as a cue. That is, participants used a "friend?" strategy, but neglected to use the intuitively more appropriate "foe?" strategy for detecting untrustworthy faces, even when explicitly told that there were dangerous individuals amongst the faces. This may have to do with the base rates of truly dangerous individuals in our daily lives. Expending cognitive resources in a constant search for untrustworthy individuals likely represents overkill on the part of the observer. Instead, simply confirming the kindness (and by extension, trustworthiness) of strangers may simply represent the most efficient means of discerning who poses an immediate threat to us. This would be consistent with our results, which show that judgments for trustworthy faces are more accurate and consistent than judgments for untrustworthy faces. Thus, trustworthy faces may have been rapidly sorted as such through the use of a kindness-based intuitive discrimination strategy, whereas untrustworthy faces may have been classified as ambiguous because of their relative lack in perceived kindness, thereby being less accurately sorted.

To our knowledge, the accuracy of trustworthiness judgments has not previously been examined. Such judgments may have evolved to be highly accurate and/or now could simply involve guessing. Overall, participants judged the trustworthiness of the targets above chance even after seeing their faces for only 100 ms. These results are consistent with past findings that people can successfully infer certain traits on the basis of photographs (Borkenau & Liebler, 1993; Roney et al., 2006; Shevlin, Walker, Davies, Banyard, & Lewis, 2002). However, although this suggests that intuitive judgments have some validity, the judgments were far from perfect; the mean accuracy rates were only 55.8% and 57.9% at Times 1 and 2, respectively. In other words, judgments were inaccurate more than 40% of the time. Furthermore, at Time 1, the criminals were as likely to be judged trustworthy as they were to be judged untrustworthy (48.8% accuracy). These results suggest that initial impressions of trustworthiness may be less accurate for potentially dangerous targets. Although confidence is formed rapidly and endures with time, the current findings suggest that its relationship with accuracy is a tenuous one. Participants were less accurate in their initial judgments of untrustworthy targets relative to trustworthy targets but were similarly confident in both judgment types.

It is important to note that a small "trustworthy" response bias for the initial judgments may have contributed to the higher accuracy for initial judgments of trustworthy targets. This finding

is in line with previous research in the area of deception suggesting that people are more likely to judge others as truthful in the absence of background information about credibility (e.g., O'Sullivan et al., 1988). When participants were informed of the nature of the two groups, the response bias disappeared, but above-chance judgments still occurred; accuracy in judging the untrustworthy targets increased, and accuracy in judging the trustworthy targets remained stable. Thus, the initial response bias cannot fully account for the findings that judgments were above chance at both Times 1 and 2. Why did the response bias disappear in the final set of judgments? First, participants may have incorporated knowledge of the base rates into their decision making. Second, they may have become more cautious or sceptical in judging the targets after they were provided with such information. Participants' new knowledge that they might encounter potentially dangerous faces may have facilitated a more critical assessment and put them on higher alert. The finding that the initial judgments of trustworthy targets were more likely to be retained (stay the same) in the final judgments relative to untrustworthy targets was not entirely consistent with the idea that accurate interpersonal assessment can be impaired by the assimilation of new information concerning the target to match the initial judgments (O'Sullivan, 2003). The results suggested that this assimilation process did not occur for the untrustworthy targets specifically.

What was it about the target faces that led to particular inferences of their level of trustworthiness? At the time of the initial judgment, targets rated as trustworthy had been given lower ratings of aggressiveness and higher ratings of kindness than targets rated as untrustworthy, supporting the results of Richell et al. (2005). Because previous research has suggested that humans have evolved to detect threatening, angry faces rapidly and accurately (e.g., Williams & Mattingley, 2006), it may be the case that we are "programmed" to rapidly perceive threatening facial expressions and make concordant inferences of trustworthiness. Finally, it has been suggested that traits such as hostility may actually influence facial appearance because the vascular and muscular properties of the face will be affected by the repeated expression (Malatesta, Fiore, & Messina, 1987; Zajonc, 1985).

This study had limitations to consider in interpreting the results. First, participants were mostly female undergraduate students, limiting the generalizability of the findings. Moreover, only male faces were presented as targets; it would be useful to examine trustworthiness judgments of female targets in future research. Also, situational factors may have biased groups toward having either low or high perceived trustworthiness. For instance, displays of particular facial emotions (e.g., anger) are associated with particularly low ratings of perceived trustworthiness (e.g., Richell et al., 2005). Although efforts were made to equate emotional expressions across groups, it might be argued that the fugitive faces were characterised by subtle anger expressions or the award winners by subtle smiles, as images typically came from mugshots and press photos, respectively. However, we believe that this limitation was at least partially controlled for by using kindness and aggressiveness discrimination in our analyses, which, in turn, provided some interesting insights into how rapid impressions of trustworthiness are formed.

The results of the current study suggest that intuition plays a small but facilitative role in judging the trustworthiness of a stranger on the basis of facial appearance. Specific cues, such as

perceived kindness or aggressiveness, differed between groups and were either used effectively, as was the case for kindness, or counterproductively, as was the case for aggressiveness, to form first impressions of trustworthiness that are above chance in accuracy. These findings have important implications for future credibility assessment research, demonstrating that intuition, although often misused by deception detectors (e.g., Porter et al., 2000), may confer some benefit in particular situations. Future research might explore the role of specific characteristics or facial regions involved in making intuitive trustworthiness judgments. For example, the eye region is automatically attended to when first viewing a face, a process controlled by a region of the brain associated with the perception of both facial emotions and trustworthiness, the amygdala (see Adolphs et al., 2005). This region, or others, may play a preferential role informing our initial judgments of trustworthiness.

We conclude that intuition lends a small advantage when making assessments of trustworthiness based on facial appearance, but errors are common. However, the knowledge that some targets to be encountered are untrustworthy serves to increase the accuracy of identifying such targets. Finally, extending exposure time beyond a fleeting glance does not change or improve judgments of trustworthiness. Overall, our results suggest that the face is a rather opaque window to the soul.

Résumé

Bien que les jugements de confiance portés à la vue d'un visage étranger se forment rapidement (Willis & Todorov, 2006), leur exactitude est inconnue. Nous avons étudié l'exactitude des jugements de confiance portés à la vue du visage de personnes appartenant à deux groupes différents, auxquels on accorde une confiance opposée (récipiendaires d'un Prix Nobel de la paix d'un prix humanitaire par opp. aux criminels les plus recherchés aux États-Unis). Les participants ont visionné 34 visages présentés pendant un 1/10 de seconde ou 30 secondes et ils ont évalué la confiance qu'ils leur accordaient. Par la suite, les participants ont été informés de la nature de chacun des groupes et ont associé le groupe avec chacun des visages. Les jugements portés à la suite d'une très brève exposition étaient similaires sur le plan de l'exactitude et de la confiance à ceux formés à la suite d'une longue exposition. Toutefois, les jugements portés à la vue d'un visage qui n'inspiraient pas confiance aux participants (visages d'un criminel) étaient moins précis ($M = 48,8\%$) que ceux à qui les participants accordaient leur confiance ($M = 62,7\%$). L'exactitude du jugement était aléatoire pour les cibles de confiance de la première phase seulement et légèrement aléatoire pour les deux types de cibles visionnées à la deuxième phase. Les participants se fiaient à la bonté et à l'agressivité apparentes pour orienter les décisions prises rapidement. Ainsi, l'intuition joue un faible rôle de facilitation lors de l'observation d'un visage et les erreurs sont courantes.

Mots-clés : visage, fiabilité, intuition, crédibilité

References

- Adolphs, R. (2002). Trust in the brain. *Nature Neuroscience*, 5, 8–9.
 Adolphs, R., Gosselin, F., Buchanan, T., Tranel, D., Schyns, P., &

- Damasio, A. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*, 68–72.
- Ambady, N., Krabbenhoft, M. A., & Hogan, D. (2006). The 30-sec sale: Using thin-slice judgements to evaluate sales effectiveness. *Journal of Consumer Psychology*, *16*, 4–13.
- Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*, 269–278.
- Borkenau, P., & Liebler, A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology*, *65*, 546–553.
- Bull, R. (2006). Detecting lies and deceit: The psychology of lying and the implications for professional practice. *Journal of Community & Applied Social Psychology*, *16*, 166–167.
- Bull, R., & Green, J. (1980). The relationship between physical appearance and criminality. *Medicine, Science and the Law*, *20*, 79–83.
- Bull, R., & Vine, M. (July 2003). *Attractive people tell the truth: Can you believe it?* Poster presented at the Annual Conference of the European Association of Psychology and Law, Edinburgh.
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind and the generation of culture* (pp. 163–228). Oxford, England: Oxford University Press.
- Darwin, C. (1872). *The expression of the emotions in man and animals*. Chicago: University of Chicago Press.
- Dumas, R., & Teste, B. (2006). The influence of criminal facial stereotypes on juridic judgments. *Swiss Journal of Psychology*, *65*, 237–244.
- Hare, R. D. (2003). *Manual for the Revised Psychopathy Checklist* (2nd ed.). Toronto: Multi-Health Systems.
- Hellström, A., & Tekle, J. (1994). Person perception through facial photographs: Effects of glasses, hair, and beard on judgments of occupation and personal qualities. *European Journal of Social Psychology*, *24*, 693–705.
- Lipton, E. (2006, August 17). Threats and responses: Screening; faces, too, are searched as U.S. airports try to spot terrorists. *New York Times*. Retrieved October 20, 2006, from <http://www.nytimes.com/2006/08/17/Washington/17screeners.html>
- Little, A., & Perrett, D. (2007). Using composite images to assess accuracy in personality attribution to faces. *British Journal of Psychology*, *98*, 111–126.
- Macrae, C. N., & Shepherd, J. (1989). Do criminal stereotypes mediate juridic judgments? *British Journal of Social Psychology*, *28*, 189–191.
- Malatesta, C. Z., Fiore, M. J., & Messina, J. J. (1987). Affect, personality, and facial expressive characteristics of older people. *Psychology and Aging*, *2*, 64–69.
- Martelli, M., Majib, J. M., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision*, *5*, 58–70.
- Meissner, C., & Kassin, S. (2002). He's guilty! Investigator bias in judgments of truth and deception. *Law and Human Behavior*, *26*, 469–480.
- O'Sullivan, M. (2003). The fundamental attribution error in detecting deception: The boy-who-cried-wolf effect. *Personality and Social Psychology Bulletin*, *29*, 1316–1327.
- O'Sullivan, M., Ekman, P., & Friesen, W. V. (1988). The effect of comparisons on detecting deceit. *Journal of Nonverbal Behavior*, *12*, 203–215.
- Penton-Voak, I., Pound, N., Little, A., & Perrett, D. (2006). Personality judgements from natural and composite facial images: More evidence for a “kernel of truth” on social perception. *Social Cognition*, *24*, 607–640.
- Pessoa, L., Japee, S., Sturman, D., & Ungerleider, L. G. (2005). Target visibility and visual awareness modulate amygdala responses to fearful faces. *Cerebral Cortex*, *16*, 366–375.
- Petrides, K. V., & Furnham, A. (2003). Trait emotional intelligence: Behavioural validation in two studies of emotion recognition and reactivity to mood induction. *European Journal of Personality*, *17*, 39–57.
- Petrides, K. V., & Furnham, A. (2006). The role of trait emotional intelligence in a gender-specific model of organizational variables. *Journal of Applied Social Psychology*, *36*, 552–569.
- Porter, S., & ten Brinke, L. (in press). Dangerous decisions: A theoretical framework for understanding how judges assess credibility in the courtroom. *Legal and Criminological Psychology*.
- Porter, S., & Woodworth, M. (2007). “I’m sorry I did it . . . but he started it”: A comparison of the official and self-reported homicide descriptions of psychopaths and non-psychopaths. *Law and Human Behavior*, *31*, 91–107.
- Porter, S., Woodworth, M., & Birt, A. R. (2000). Truth, lies, and videotape: An investigation of the ability of federal parole officers to detect deception. *Law and Human Behavior*, *24*, 643–658.
- R. v. B. (K. G.) 1993 CanLII 116 (S.C.C.), (1993), 79 C.C.C. (3rd) 257 (S.C.C.).
- Richell, R. A., Mitchell, D. G. V., Peschardt, K. S., Winston, J. S., Leonard, A., & Dolan, R. J., et al. (2005). Trust and distrust: The perception of trustworthiness of faces in psychopathic and non-psychopathic offenders. *Personality and Individual Differences*, *38*, 1735–1744.
- Roney, J. R., Hanson, K. N., Durante, K. M., & Maestripieri, D. (2006). Reading men’s faces: Women’s mate attractiveness judgments track men’s testosterone and interest in infants. *Proceedings of the Royal Society of London B*, *273*, 2169–2175.
- Shevlin, M., Walker, S., Davies, M., Banyard, P., & Lewis, C. (2002). Can you judge a book by its cover? Evidence of self-stranger agreement on personality at zero acquaintance. *Personality and Individual Differences*, *35*, 1373–1383.
- Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005, June 10). Inferences of competence from faces predict election outcomes. *Science*, *308*, 1623–1626.
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Chichester, England: Wiley
- Williams, M. A., & Mattingley, J. B. (2006). Do angry men get noticed? *Current Biology*, *16*, 402–404.
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*, 592–598.
- Zajonc, R. B. (1985, April 5). Emotion and facial efference. *Science*, *228*, 15–21.
- Zebrowitz, L. A., Voinescu, L., & Collins, M. A. (1996). “Wide-eyed” and “crooked-faced”: Determinants of perceived and real honesty across the life span. *Personality and Social Psychology Bulletin*, *22*, 1258–1269.

Received December 13, 2007

Revision received March 5, 2008

Accepted March 18, 2008 ■