Incidental Learning of Trust: Examining the Role of Emotion and Visuomotor Fluency

James W. A. Strachan and Alexander J. Kirkham
University of York

Luis R. Manssuer
Bangor University

Steven P. Tipper
University of York

Eye gaze is a powerful directional cue that automatically evokes joint attention states. Even when faces are ignored, there is incidental learning of the reliability of the gaze cueing of another person, such that people who look away from targets are judged less trustworthy. In a series of experiments, we demonstrated further properties of the incidental learning of trust from gaze direction. First, the emotion of the face, whether neutral or smiling, influenced the pattern of trust learning. Second, the effect was specific to judgments of trust; reliability of gaze direction did not influence other emotional judgments of a person, such as liking. And third, visuomotor fluency was not sufficient for learning of trust, whether or not the face served as a target or as a distractor. Taken together, incidental learning of trust is influenced by facial emotion, it is a specific effect that does not generalize to other emotional assessments, and it is not determined solely by processing fluency.

Keywords: gaze cueing, task switching, visuomotor fluency, emotion, trustworthiness, likability

Supplemental materials: http://dx.doi.org/10.1037/xlm0000270.supp

There is a scene in the 2006 James Bond film Casino Royale where Daniel Craig’s Bond is struggling with another man over a knife. During the fight, Bond hesitates and stares over the other man’s shoulder, distracting his opponent temporarily and granting him the upper hand. For most people, fights to the death are thankfully uncommon, but this misdirection of one’s attention is a common trope that we all easily recognize. Redirection of our attention based on somebody else’s can be perfectly innocent, as when someone genuinely spots something that demands their and perhaps your own attention, or it can be the result of a calculated deception, as in the case of sleight of hand magic tricks. Interpreting such events in terms of their social implications therefore reflects a key element of social cognition.

Hence eye gaze is a powerful communicative tool. It can be used to reflexively redirect another’s attention toward or away from a particular object or location (Driver et al., 1999; Friesen & Kingston, 1998; Frischen & Tipper, 2004). Due to its dual ability to facilitate attentional processing or to misdirect it and incur a cost to processing fluency, people can also infer higher order information from this; objects that are looked at tend to be liked more (Bayliss, Paul, Cannon, & Tipper, 2006; Capozzi, Bayliss, Elena, & Becchio, 2015; Manera, Elena, Bayliss, & Becchio, 2014; Ulloa, Marchetti, Taffou, & George, 2015) and those who correctly cue an object location are chosen as more trustworthy than those who mislead (Bayliss & Tipper, 2006; Bayliss, Griffiths, & Tipper, 2009; Manssuer, Pawling, Hayes, & Tipper, 2015; Rogers et al., 2014).

In the initial investigation of incidental learning of trust from gaze behavior, Bayliss and Tipper (2006) used a gaze-cueing paradigm with a group of paired faces. One face of each pair would always look toward a target object that participants had to identify as a kitchen or a garage item (valid cue), while the other would always look away from the target (invalid cue). At the end of the experiment, participants were shown each face in the pair and asked to select which they thought was more trustworthy.

This effect has been shown using this two-alternative forced choice (2AFC) rating procedure. However, this does not explain how these changes come about—whether this effect is driven by an increase in trustworthiness for valid faces, a decrease for invalid faces, or a bidirectional mix of the two. To further investigate the specific nature of changes in trust ratings, we employed two scalar ratings of trustworthiness in the current studies, one at the beginning and one at the end of the experiment, to track changes in trustworthiness for both valid and invalid faces (cf., Manssuer, Roberts, & Tipper, 2015). This more sensitive measure provides the ideal approach to further investigate key boundary conditions.
for the understanding of the processes mediating incidental learning of trust.

One outstanding issue where this new measure may be beneficial concerns the role of facial emotion. Bayliss et al. (2009) found that gaze-contingent trust effects appear to rely on a positive social context, because they found no trust effects when the faces expressed anger and a reliable effect only when the faces smiled. However, the neutral expression condition was somewhat ambiguous, as participants were only slightly more likely to select the valid face as the more trustworthy of a matched pair in a 2AFC paradigm. This previous work using forced choice between valid and invalid cueing individuals creates a somewhat blunt measure; we can see that valid faces are preferred over invalid, but we do not know if this is because valid faces become more trustworthy, invalid become less, or some combination of the two. We also do not know how exactly the emotional expression of a face might change the pattern of results using forced-choice measures of trustworthiness. It still could be the case that neutral faces are not sufficient to elicit learning of trust—there is a wealth of evidence that smiling faces are treated differently from neutral faces in various social interactions, both when measured by trustworthiness judgments (Hehman, Flake, & Freeman, 2015) and by more implicit measures (Wang & Hamilton, 2014)—or it may be that we can detect trust learning with neutral faces using this new, more sensitive measure.

Therefore we examined the role of emotion in the incidental learning of trust in conditions where faces express neutral emotions (Experiment 1) and when they smile (Experiment 2). Doing so would enable us to unequivocally identify whether incidental learning of trust from gaze cueing can be detected when faces express neutral emotions. Additionally, and more importantly, this would allow us to assess whether the pattern of learning (whether valid faces increase in trust and invalid faces decline in trustworthiness, or whether the effects are unidirectional) is the same for both neutral and positive emotions.

In addition, we investigated whether incidental learning of eye-gaze patterns is specific to judgments of trust, or generalizes to other emotional assessments, such as liking a person. One might assume that trust and liking will be closely related: If we trust someone, we are more likely to like them. Indeed, the two are often conflated as aspects of warmth in dual-dimension theories of social cognition (e.g., Fiske, Cuddy, & Glick, 2007). However, subtle behaviors that can be used to deceive others, such as gaze shifts, could have quite specific effects on trust. For example, whether to invest money with another person is influenced by incidental learning of patterns of gaze shifts, as is the decision to be altruistic while computing the likelihood that such an act will be reciprocated in the future (Rogers et al., 2014). Such decisions might not be affected by general feelings of liking, for example, we may trust a lawyer to do his or her utmost to preserve our freedom, but we may not like that lawyer on a personal level; the two feelings are distinct, and can be separated. To our knowledge, there is little previous work directly addressing whether trust and liking are functionally similar in this way, so, in Experiment 3, we replaced the trustworthiness ratings of Experiment 1 with likability ratings, to see if this incidental learning is specific to trust or if there is a broader affective spillover into other social judgments.

The final issue concerns the role of visuomotor fluency in the learning of trust. In the gaze-cueing procedure, there are two aspects that might mediate the learning of trust. One is the behavior of another person. As noted, gaze can be used to help or deceive another person. That is, looking toward interesting and desirable objects to facilitate a conspecific’s behavior or looking away from desirable objects to mislead. The second is the visuomotor fluency experienced during gaze cueing. That is, responses are faster on valid trials where gaze directs the other person’s attention to the location where a target will appear. Previous work has shown that facilitating both perceptual (e.g., Reber, Winkielman, & Schwarz, 1998) and motor (e.g., Hayes, Paul, Beuger, & Tipper, 2008) performance increases preference and liking of images and objects.

Therefore we investigated the incidental learning of trust in a task where increased processing fluency is associated with some faces and impaired processing fluency is associated with other faces in a similar way to gaze-cueing studies. However, there are no face behaviors, such as gaze shifts, that might be associated with deception. To this end, we developed two new task-switching procedures in Experiments 4 and 5, and consistently associated some face identities with fluent, fast, and accurate processing and other faces with impaired, slow, and error prone processing. In Experiment 4, we developed a task-switching procedure where the faces were now the targets of participants’ decisions, and, in Experiment 5 we used a design where the faces remained background distractors. If visuomotor fluency is the key driver for the learning of trust, then effects should be detected in one or both of these new tasks. On the other hand, if cues to deception such as eye gaze are necessary, then no learning of trust should be detected.

To briefly preview our findings: We found that face emotion does influence the learning of trust from gaze behavior, but in a selective manner. That is, when faces express neutral emotions, there was a decline in trust for invalid faces that look away from targets. No such effect was observed with valid faces, except when these faces were seen smiling. When examining the generalizability of the trust effect, somewhat counterintuitively, we found no effects when assessing liking of another person. And, finally, in both task-switching procedures, where the faces were targets, and where they were distractors, there were no changes in trust assessments, suggesting that changes in visuomotor fluency were not sufficient to generate learning of trust.

Experiment 1

We reexamined whether incidental learning of trust can be detected with faces expressing neutral emotion, and, if such an effect is detected, what pattern of changes in trust are revealed?

Method

Participants. A total of 24 participants (18 female) volunteered for the study in return for payment or course credit. All were students of the University of York, and had a mean age of 19.96 years. All participants in all experiments described in this study provided written consent and the research was given ethics approval by the Ethics Committee of the Department of Psychology at University of York.

Stimuli. Target stimuli for the object classification task were the kitchen and garage object images used in Bayliss and Tipper (2006). There were 13 unique objects in each group (kitchen/
Design and procedure. Participants were told that they would be asked to perform an object categorization task on images of objects that appeared on the left or the right side of the screen, and to respond with whether the images were garage or kitchen objects. They were also told that the central face images were irrelevant and to be ignored. Before the experiment, participants were allowed to study printed versions of the kitchen/garage images, in order to familiarize themselves. This was done first to ensure that participants had the knowledge of what each object was, and second to make sure that early responses from the first trial block were not confounded by uncertainty about object categories of the targets.

Each trial began with a 600-ms fixation cross in the center of the screen, which was then replaced by a face showing a direct gaze for 1,500 ms. The face then shifted gaze either to the left or the right for 500 ms before the target stimulus appeared on either the same (valid) or opposite (invalid; see Figure 1a) side of the gaze direction. The target stimulus remained either until the participant’s response was logged or until 2,500 ms had passed, following which participants received feedback from an error tone that would sound if an incorrect response was logged. The face then shifted back to direct gaze for another 1,000 ms. A blank screen followed for 500 ms before the next trial began. The trial structure is shown in Figure 1b.

The object categorization responses were the H key and the space bar of a keyboard, chosen because the H key appears directly above the space bar on QWERTY keyboards and this direction was orthogonal to the possible location of the target. Participants were instructed to respond with their index finger on the H key and thumb on the space bar. For half the participants, H represented kitchen objects, while for the other half it represented garage objects.

In total, there were five blocks of 32 trials each, and each face appeared twice in each block, once gazing left and once right (10 times in total across the experiment; five left, five right). The order of faces was randomized, as was the order of target objects, the side that the target appeared, and the order of valid and invalid trials.

At the beginning and end of the experiment, participants rated the original unmanipulated face images used to generate the gaze-cueing stimuli. Participants were shown a calibration slide where they clicked in the center to start, and then the face images were presented for 1,000 ms. Participants were then instructed to click along an uninterrupted scale at a point that conformed to how trustworthy they felt the person was. The scale recorded response clicks between −100 and +100, calculated by the distance from the center of the scale—responses to the left of the center of the scale were coded as negative, while those to the right were coded as positive (these were indicated on the screen with a − and + sign at either end of the scale). Identities were presented in a randomized order.

After the experiment, we asked all participants what they thought the nature of the experiment had been and if they had picked up on the experimental manipulation. While some participants did demonstrate suspicion of the manipulation in this and later gaze-cueing experiments, it was rare for any participant to spontaneously describe the pattern of eye gaze. To offer a qualitative interpretation, many answers were given hesitantly and made it appear that participants were thinking back over the structure of the experiment and inferring from that in order to generate an answer, rather than as a result of their own in-the-moment intuition during gaze cueing.

Data analysis. Before data were analyzed, participants’ responses were filtered to remove all error trials (where participants reported the incorrect answer) and reaction time (RT) outliers—RTs below 250 ms (too short to process the stimuli) and above 2,500 ms (indicating that participants had not given a response in the allotted time). The number of remaining trials was then compared with the original number of trials to check that all participants retained at least 70% of their total trials and had not scored below 70% total correct on any one condition. Mean RTs and percentage accuracy scores were calculated for each participant for both valid and invalid trials for each block separately. RTs and accuracy rates for each block were compared in separate 2 × 5 (Validity × Block) repeated-measures analyses of variance (ANOVAs).

Average trustworthiness ratings were calculated for each participant both at the beginning (pre) and end (post) of the experiment for both valid and invalid faces, and these scores were analyzed in a 2 × 2 (Time × Validity) repeated-measures ANOVA.

1 Although means and standard deviations were not retrieved from Östner and Todorov’s (2008) material, we can validate our assumption that these groups of faces were close to neutral by examining the preratings assigned to them in the five experiments presented here. Because the preratings occurred before any participants had a chance to experience the faces within an experimental context, any differences could only be explained by physical cues to trustworthiness, and the combined power of these five experiments would be sufficient to detect this. We explored this and found that the preratings for faces in one group (M = 6.82, SD = 13.04) did not significantly differ from the other group (M = 0.85, SD = 6.82), t(14) = −0.68, p = .506.
Results and Discussion

Gaze cueing. The RT results of Experiment 1 are shown in Figure 2a. Over the course of the five blocks, RTs were slower in valid than in invalid trials. A $2 \times 5$ ANOVA found a main effect of validity, $F(1, 23) = 47.94, p < .001$, $\eta^2_p = 0.22$, and a main effect of block where responses were faster in later blocks than earlier, using the Greenhouse–Geisser correction for violation of sphericity assumption, $F(1.92, 44.16) = 21.51$, $p < .001$, $\eta^2_p = 0.13$, but no interaction, $F(4, 92) = 1.14, p = .345$, $\eta^2_p = 0.00$.

A similar $2 \times 5$ ANOVA using accuracy scores (coded as percent correct in each trial type in each of the five blocks; see Table 1) found only a main effect of block, $F(4, 92) = 3.36, p = .013$, $\eta^2_p = 0.05$, as participants generally committed more errors at the beginning of the experiment than at the end, but there was no overall effect of cueing validity on errors, $F(1, 23) = 0.15, p = .700$, $\eta^2_p = 0.00$, and no Validity $\times$ Block interaction, $F(4, 92) = 1.46, p = .220$, $\eta^2_p = 0.01$. These results suggested that the attention-cueing effect emerged primarily in RT measures rather than error rates, and that it remained stable over time.

Trustworthiness ratings. A repeated-measures ANOVA with rating time (pre- and postexperiment) and cueing validity of faces (valid and invalid) as within-subjects factors found a significant
overall effect of time where ratings were lower after the experiment than before, $F(1, 23) = 10.49, p = .004, \eta^2_p = 0.08$, and one of cueing validity where ratings were lower for invalid faces than valid, $F(1, 23) = 7.23, p = .013, \eta^2_p = 0.13$. There was also a significant Time of Rating $\times$ Cueing Validity interaction, $F(1, 23) = 7.19, p = .013, \eta^2_p = 0.09$. Figure 3a shows how the trustworthiness ratings for each group changed during the experiment. It is clear that there were no changes in trust for faces that consistently looked toward targets, Valid: $t(23) = 0.07, p = .946$, whereas there was a significant decline in trustworthiness for faces that consistently looked away from targets, Invalid: $t(23) = -4.19, p < .001$.

The results of Experiment 1 demonstrated that the trust effect can be obtained with neutral faces, where previously such effects...
were not clearly demonstrated (e.g., Bayliss et al., 2009). It is possible that the pre- and postexperiment assessments that allow measures of change in trust for each face, and the use of an unmarked analog scale, were more sensitive measures than the forced-choice decision used in previous work. Finally, it is noteworthy that there was an asymmetry in the effect that only invalid faces declined in trustworthiness, while valid faces did not change in trust ratings. Experiment 2 explored whether this would hold true when faces expressed positive emotions.

**Experiment 2**

This experiment sought to explore how emotion affects this incidental learning of trust. It replicated all details of Experiment 1, but used smiling rather than neutral face images as the cuing and rating stimuli. Note that Bayliss et al. (2009) demonstrated significant learning of trust when the faces expressed positive emotions with a smile. However, when the faces expressed a neutral emotion, the same pattern of trust was observed, but it was of marginal significance. Experiment 1 showed that it is possible to detect significant learning of trust when faces are neutral, however, the effect was asymmetrical, because invalid faces declined in trust and valid faces did not change. Whether faces expressing positive emotions would produce this same pattern was the key question for Experiment 2.

**Method**

**Participants.** Twenty-four participants (21 female; \( M_{\text{age}} = 20.46 \) years) volunteered for this study in return for course credit or payment.

**Stimuli, design, and procedure.** This experiment was identical to Experiment 1 in every way except that the KDEF faces used were frontal-view smiling faces rather than neutral faces both during the gaze-cuing portion of the experiment and at both pre- and posttrustworthiness rating presentations. All other details were identical.

**Data analysis.** RT filters and analysis of RTs, error rates, and trustworthiness ratings data were identical to those in Experiment 1.

**Results and Discussion**

**Gaze cueing.** The results of Experiment 2 are shown in Figure 2b. Over the course of the five blocks, RTs were faster for valid than for invalid trials. A \( 2 \times 5 \) ANOVA found a main effect of validity, \( F(1, 23) = 25.58, p < .001, \eta^2 = 0.25 \), and a main effect of block, Greenhouse–Geisser corrected: \( F(8.6, 42.72) = 31.33, p < .001, \eta^2 = 0.25 \), but no interaction, Greenhouse–Geisser corrected: \( F(4, 92) = 1.22, p = .306, \eta^2 = 0.00 \). A \( 2 \times 5 \) (Validity \( \times \) Block) ANOVA on the accuracy rates (see Table 1) found the main effect of block approached but was not significant, \( F(4, 92) = 2.07, p = .091, \eta^2 = 0.05 \), and there was no effect of validity, \( F(1, 23) = 0.09, p = .769, \eta^2 = 0.00 \), or interaction between the two, \( F(4, 92) = 1.27, p = .287, \eta^2 = 0.01 \).

**Trustworthiness ratings.** The changes in trustworthiness ratings for the faces in Experiment 2 are shown in Figure 3b. A repeated-measures ANOVA with rating time (pre- and postexperiment) and cueing validity of faces (valid and invalid) as within-subjects factors found no overall effect of time, \( F(1, 23) = 0.02, p = .902, \eta^2 = 0.00 \), but did find a significant effect of cueing validity where valid faces were rated higher than invalid faces, \( F(1, 23) = 5.48, p = .028, \eta^2 = 0.07 \), as well as a significant Time of Rating \( \times \) Cueing Validity interaction, \( F(1, 23) = 12.08, p = .002, \eta^2 = 0.01 \). As in Experiment 1, there was a decline in trustworthiness ratings for invalid faces that looked away from targets, which in this experiment approached significance, invalid: \( t(23) = -1.78, p = .088 \), but we also detected a significant increase in trustworthiness for valid faces that looked toward targets, valid: \( t(23) = 2.20, p = .038 \).

To explore the contrasts between Experiments 1 and 2, we assessed post hoc tests examining the change in trust for valid faces and invalid faces. Due to violations of the normality assumption, separate Mann–Whitney \( U \) tests were performed on the data. When examining the change in trust for valid faces that consistently looked toward targets, increases in trust were greater when the faces were smiling than when neutral \( (U = 181, p = .028) \). In contrast, whether faces were smiling or neutral had no effect on the decline in trust ratings for the invalid faces that looked away from targets \( (U = 282, p = .909) \).2

As noted, Bayliss et al. (2009) observed that smiling faces produced more robust trust effects than neutral faces. However, this previous work required forced choice between previously valid and invalid faces. Such forced-choice measures cannot identify whether the effects are specific to valid faces increasing in trust, invalid decline in trust, or both. The current approach enabled a more detailed analysis where change in trust ratings can identify the specific patterns of trust effects. Experiment 2 showed that the more robust trust effects were specifically due to increases in trust of valid faces only when they were smiling. Conversely, a decrease in trustworthiness for invalid faces appeared to be the

---

2For the sake of completeness, we also compared the results of Experiments 1 and 2 in a mixed \( 2 \times 2 \) (Cueing Validity, Within Experiment, Between ANOVA, with change in trustworthiness (calculated as preexperiment ratings subtracted from postexperiment ratings) as the dependent variable, and with planned contrasts between valid and invalid faces across experiments. We found a significant overall effect of validity, \( F(1, 46) = 19.21, p = .001, \eta^2 = 0.25 \), but none of experiment, \( F(1, 46) = 1.56, p = .218, \eta^2 = 0.02 \), and no interaction between the two, \( F(1, 46) = 0.77, p = .386, \eta^2 = 0.01 \). Planned comparisons found that the difference between neutral and smiling valid faces approached significance \( (p = .079) \), but there was no such difference between neutral and smiling invalid faces \( (p = .624) \).
stable, key feature of this effect. By comparing trust learning in response to neutral faces with that of smiling faces, we have provided the first evidence that this incidental learning is asymmetrical (for a detailed description of the possible implications and mechanisms underlying this asymmetricality, see the General Discussion).

Based on previous research (Bayliss et al., 2009; Pecchinenda, Pes, Ferlazzo, & Zoccolotti, 2008), we might expect that social expectations driven by emotional expression (i.e., that smiling faces are likely to be helpful) might influence the incidental learning of trustworthiness. That is, valid smiling faces confirm the expectations and so lead to a stronger increase in trust, while
invalid smiling faces violate the expectation and so lead to a decrease in trust. In contrast, a neutral stare for social primates can be perceived as a threat and hence the detection of negative events, such as potential deception by invalid faces that consistently look away from targets, takes precedence and there is little learning of the positive events such as joint attention produced by valid faces.

**Experiment 3**

Experiments 1 and 2 confirmed that even when a face is irrelevant and could be ignored while focusing on the main task of peripheral target classification, there was learning of the gaze patterns of another person. That is, an association between face identity and reliability of gaze direction was learned and retrieved when reencountering faces at a later time. Thus, incidental learning of gaze direction subsequently changed how much another person is trusted, and this was affected to some extent by the emotional expression of the person.

The next studies further investigated the properties of this incidental learning process. Experiment 3 explored whether this impression is specific to trust or simply reflects a broader valence impression of the face. Therefore, this experiment was identical to Experiment 1, except for one very minor change: that is, the question that participants were asked was changed from one of “How trustworthy do you think this person is?” to “How likeable do you think this person is?”

**Method**

**Participants.** Twenty-six participants (21 female; M_age = 18.99 years) volunteered for this study in return for course credit or payment. One participant’s data were not collected due to a computer malfunction, and one participant had to be removed after RT filters were applied, so the final sample available for analysis was 24.

**Stimuli, design, and procedure.** This experiment was identical to Experiment 1 except that, at the beginning and the end of the experiment, participants were asked, “How likeable do you think this person is?” rather than “How trustworthy do you think this person is?” All mention of trustworthiness on consent forms was changed to “likeable” (dependent on context).

**Data analysis.** RT filters were applied in the same way as in Experiments 1 and 2, and, in this experiment, one participant had to be removed for retaining less than 70% of his or her original trials.

Average likability ratings were calculated for each participant both at the beginning (pre) and the end (post) of the experiment for both valid and invalid faces, and these scores were analyzed in a 2 × 2 (Time × Validity) repeated-measures ANOVA.

**Results and Discussion**

**Gaze cueing.** The results of Experiment 3 are shown in Figure 2c. Over the course of the five blocks, RTs were faster for valid than for invalid trials. A 2 × 5 ANOVA found a main effect of validity, F(1, 23) = 14.45, p = .001, η^2_p = 0.23, but no interaction. Greenhouse–Geisser corrected: F(2.44, 56.13) = 1.60, p = .377, η^2_G = 0.00.

A 2 × 5 ANOVA on accuracy rates (see Table 1) found no main effects of block or validity, or any interaction between the two.

**Likability ratings.** The changes in likability ratings for the faces in Experiment 3 are shown in Figure 3c. A repeated-measures ANOVA with rating time (pre- and postexperiment) and cueing validity of faces (valid and invalid) as within-subjects factors found no overall effect of time, F(1, 23) = 0.20, p = .655, η^2_p = 0.00, or of cueing validity, F(1, 23) = 0.04, p = .837, η^2_p = 0.00, and no significant interaction between the two, F(1, 23) = 2.26, p = .147, η^2_p = 0.03.

To explore the contrasts between Experiments 1 and 3, we assessed post hoc tests examining the change in ratings for valid faces and invalid faces. Due to violations of the normality assumption, separate Mann–Whitney U tests were performed on the data. The change in ratings (trust vs. liking) for invalid faces that consistently looked away from targets showed a significantly greater decrease in trust ratings than likability ratings (U = 166.5, p = .013), but no such differences between ratings emerged for valid faces (U = 267, p = .673).

The results of Experiment 3 were quite surprising. This experiment was identical to Experiment 1 except for a single word change (from “trustworthy” to “likeable”) in the ratings task. Although the RT gaze-cueing effects were the same in Experiments 1 and 3, this gaze cueing had no effect on how likable a person was perceived to be. It was somewhat surprising that the predictability of gaze cueing could have such a specific effect, where individuals who consistently looked away from targets, a form of deception, were trusted less but not necessarily liked less.

Of course, we must be cautious when interpreting null results, and so in our supplemental materials we have included an additional version of this experiment (not included here because it also collected data from electroencephalography, which makes it difficult to compare with the current package) using different faces and run on a different participant pool (that experiment was conducted at Bangor University rather than at University of York). Critically, this experiment also failed to detect any effect of gaze-cueing contingencies on judgments of likability, whereas significant effects on judgments of trust have been repeatedly observed. Faces used in the supplemental materials were preselected as appearing high in happiness, which means they expressed slight smiles. Slight smiles have been effective at eliciting strong changes in trust (Manssuer, Pawling, et al., 2015; Manssuer, Roberts, & Tipper, 2015), and they are comparable to the effect found in Experiment 2 with full smiles. That we saw no learning of liking with these slightly smiling faces led us to have increased confidence in the lack of incidental learning from eye gaze on judgments of liking.

---

3 Comparing Experiments 1 and 3 with a 2 × 2 mixed ANOVA found a significant overall effect of validity, F(1, 46) = 9.44, p = .004, η^2_p = 0.10, as well as an effect of experiment, F(1, 46) = 4.94, p = .031, η^2_p = 0.08, but no interaction between the two, F(1, 46) = 2.35, p = .132, η^2_p = 0.03. Planned comparisons showed the difference between trustworthiness and likability ratings was significant for invalid faces (p < .001) but not for valid faces (p = .502).
**Experiment 4**

A key feature of the first three experiments is that trust was influenced by eye-gaze behaviour of another person. Clearly looking toward or away from relevant objects is a means of deceiving another person and initiates joint attention, which recruits reward-related neurocircuitry (Gordon, Eilbott, Feldman, Pelphrey, & Vander Wyk, 2013; Schilbach et al., 2010). However, it remains to be seen whether this effect is wholly dependent on this joint attention feature or if similar effects can be induced purely through selective disruptions of visuomotor fluency in the absence of any physical changes to the face. Previous research has shown that perceptual fluency (e.g., Reber et al., 1998) and motor fluency (e.g., Hayes et al., 2008) can influence emotional assessments of stimuli. Can impaired processing of a face with no physical changes, such as eye-gaze shifts, also influence trust judgments? To explore this, Experiment 4 was structured to assess a task-switching paradigm designed to match the gaze-cueing paradigm, where participants experienced the same disruptions of fluent processing but without any sense of shared attention with the face.

A task-switching paradigm involves participants performing two judgments of a stimulus on different trials. For example, two trials might require reporting the color of a stimulus, while the next two trials might require reporting the identity of a stimulus. These paired trials and predictable switches between tasks continue throughout an experiment. When the task changes, a visuomotor cost (slower RTs, greater probability of errors) is associated with responses on that switch trial (e.g., Monsell, 2003; Wylie & Allport, 2000; Yeung, Nystrom, Aronson, & Cohen, 2006), even when the change sequence is predictable and therefore switches can be anticipated (Kiesel et al., 2010; Rogers & Monsell, 1995). As such, Experiment 4 asked participants to make one of two judgments of face images (color or sex) where the designated task changed every other trial (a switch—repeat alternating runs paradigm).

We hypothesized that if changing visuomotor fluency evoked affective reactions (see Constable, Bayliss, Tipper, & Kritikos, 2013; Hayes et al., 2008), then creating disfluency while processing a particular face identity would reduce trust ratings. That is, throughout the experiment, particular face identities would always be presented on switch trials where RTs are slowed and errors are more likely, while other face identities would always be presented on repeat trials where RTs are fast and accurate. Hence the design matched the gaze-cueing study of Experiment 1 where a particular face identity was always presented on a valid or invalid cueing trial. However, if learning of trust is not simply based on visuomotor fluency, but rather requires specific behavior associated with deception such as eye gaze, then simply impairing processing on switch trials in the absence of any physical changes to the face should not change trust ratings.

**Method**

**Participants.** Thirty-two participants (29 female; $M_{age} = 20.97$ years) volunteered for this study in return for course credit or payment. Eight participants had to be removed after RT filters were applied, and so the final sample available for analysis was 24.

**Stimuli, design, and procedure.** Stimuli were generated from the same KDEF faces used in Experiment 1. Participants completed the same trustworthiness ratings as in Experiment 1 before and after the experiment, using full-color unaltered images. During the main portion of the experiment, however, the paradigm was changed from gaze cueing to task switching, and for this all face images were superimposed with a transparent chromatic hue in Adobe Photoshop CS6 to appear green or yellow (see Figure 4 for examples). As in Experiments 1–3, each face appeared 10 times across the experiment to each participant (twice in each block).

Participants were told that they would be asked to make one of two judgments about a face image that appeared on the screen; they would either be asked to judge the color of the image (color condition: green or yellow) or to judge the sex of the image (identity condition: male or female). Participants were told that the task they were to perform would be shown to them as a reminder before each trial, but that the task would change every other trial such that they would perform two color trials, then two identity, and so on. The act of switching between two tasks leads to responses on the first trial of the new task being slower and more error prone; this switching cost to visuomotor fluency was the critical independent variable. As such, half the identities only appeared immediately after a task switch, in the first position of the sequence (switch trial) and half appeared immediately before the switch in the second position (repeat trial). Identity and trial position was counterbalanced across participants.

During the trial, a condition cue (either “color” or “identity”) would appear on the screen for 1,000 ms, alternating every two trials) to make participants aware of the task they were performing, followed by a 500-ms fixation cross. The target image would then appear on the screen for 500 ms subtending 23.43° visual angle horizontally and 22.62° vertically, followed by a blank screen for 1,000 ms. Participants could respond at any point in this 1,500-ms window, and any response after this time window was classified as incorrect. Participant responded using the keyboard buttons Z and M, each of which corresponded to a different answer in the two tasks (i.e., Z, male and green; M, female and yellow—counterbalanced across participants).

The same RT filters were applied to the data as in Experiments 1, 2, and 3, with the difference that the 2,500-ms upper limit was shortened to 1,500 ms to reflect the timings of the experiment. Incorrect responses and responses faster than 250 ms were removed from the data and participants’ accuracy and number of trials were considered to see if they retained more than 70% of their original number of trials. This paradigm proved to be more difficult for participants than gaze cueing, because eight participants committed too many errors to be suitable for inclusion.

RT and accuracy rates were calculated for each participant for both switch and repeat trials in each of the five blocks, and they were compared in separate $2 \times 5$ repeated-measures ANOVAs. Trustworthiness ratings were calculated in the same way as in Experiment 1 with switch/repeat replacing invalid/valid as the independent measure and they were analyzed in a $2 \times 2 \times (2 \times 5)$ repeated-measures ANOVA.

**Results and Discussion**

**Task switching.** The results of Experiment 4 are shown in Figure 2d. Over the five blocks, RTs were faster to repeat than to switch trials. A $2 \times 5$ ANOVA found a main effect of trial, $F(1, 23) = 27.30, p < .001, \eta^2_g = 0.55$, and a main effect of block, Greenhouse–Geisser corrected: $F(2.64, 57.99) = 5.11, p = .001$, $\eta^2_g = 0.04$, but no interaction, $F(4, 92) = 1.74, p = .148, \eta^2_g = 0.00$. 


Figure 4. Examples of the colored stimuli used in the task-switching experiment. The original uncolored images were used during trustworthiness ratings (a), while the green (b) and yellow (c) were used in the task-switching portion. Trial sequence (d). Participants reported whether the face was colored in green or yellow or if the face was male or female, depending on a prompt before each trial. See the online article for the color version of this figure.
A 2 × 5 ANOVA on accuracy rates (see Table 1) found a main effect of trial, $F(1, 23) = 17.58, p < .001, \eta^2_g = 0.05$, with more errors on switch trials than on repeat trials, and a main effect of block, Greenhouse–Geisser corrected: $F(2.58, 59.39) = 9.33, p < .001, \eta^2_g = 0.22$, but no interaction, $F(4, 92) = 1.39, p = .243, \eta^2_g = 0.00$.

**Trustworthiness ratings.** The changes in trustworthiness ratings for the faces in Experiment 4 are shown in Figure 3d. A repeated-measures ANOVA with rating time (pre- and postexperiment) and position of face in the task sequence (switch and repeat) as within-subjects factors found a significant effect of time with more positive trustworthiness ratings after the experiment than before, $F(1, 23) = 5.02, p = .035, \eta^2_g = 0.13$, but none of face position, $F(1, 23) = 0.42, p = .521, \eta^2_g = 0.00$, and no interaction between the two, $F(1, 23) = 1.31, p = .264, \eta^2_g = 0.00$.

These results suggested that, in the absence of any physical changes to the face that led to the initiation of joint attention with participants, the change in trustworthiness judgments was not replicated. This suggested that the trust effect cannot be explained purely by disruptions of visuomotor fluency.

**Experiment 5**

Experiment 4 showed that disruptions to visuomotor fluency using a task-switching paradigm were not sufficient to change judgments of trustworthiness. However, there were a number of methodologic changes between Experiment 1 and Experiment 4, not least of which was that the latter made the faces targets of participants’ judgments, rather than distractors (we thank a reviewer for highlighting this). Literature on distractor devaluation has suggested that to-be-ignored information often shows a devaluation that to-be-attended information does not (see Raymond, 2009). It could be that we did not see any learning of trust (particularly the characteristic decrease for invalid faces evident in Experiments 1 and 2) in Experiment 4 because it was more difficult to devalue targets than distractors. If that were the case, a similar task-switching experiment to Experiment 4, but where faces are presented as background distractors, as in Experiments 1–3, might result in similar incidental learning patterns to those in Experiments 1 and 2. Experiment 5 explored this and adapted the task-switching paradigm of Experiment 4 to more closely match the object categorization task that participants completed in Experiments 1–3.

**Method**

**Participants.** Twenty-eight participants (20 female; $M_{age} = 19.07$ years) volunteered for this study in return for a mixture of course credit and payment. Four participants had to be removed after RT filters were applied, and so the final sample available for analysis was 24.

**Stimuli, design, and procedure.** This experiment closely matched the gaze-cueing experiment used in Experiment 1, but the faces no longer shifted their gaze. Instead, participants were told that they would be making one of two possible judgments on a given trial; the first was object type, where they would categorize the object as a kitchen or a garage item (as in other experiments), while the second was object color, where they would judge whether the object was blue or yellow.

Changes to the stimuli from previous experiments were the introduction of yellow-colored objects (the same objects as used in previous experiments but digitally manipulated to appear yellow instead of blue) and the fact that when faces appeared in the center of the screen we used unaltered, neutral images rather than those digitally manipulated to shift their gaze. We also introduced a task cue before each trial, to remind participants of whether they were supposed to judge the object’s type (kitchen/garage) or color (blue/yellow; see Figure 5).

The task participants completed altered every other trial in a switch/repeat task-switching procedure, as in Experiment 4, with each face presented 10 times during the experiment (twice in each of the five blocks), and a feedback tone was presented for incorrect responses. Participants were once again instructed to ignore the faces as irrelevant. Participants completed trustworthiness ratings both at the beginning and at the end of the experiment.

**Results and Discussion**

**Task switching.** The results of Experiment 5 are shown in Figure 2e. Over the course of the five blocks, RTs were faster to repeat than to switch trials. A 2 × 5 ANOVA found a main effect of trial, $F(1, 23) = 37.92, p < .001, \eta^2_g = 0.06$, but no main effect of block, Greenhouse–Geisser corrected: $F(2.53, 58.10) = 0.51, p = .727, \eta^2_g = 0.01$, and no interaction, Greenhouse–Geisser corrected: $F(2.90, 66.77) = 0.25, p = .908, \eta^2_g = 0.00$.

A 2 × 5 ANOVA on accuracy rates (see Table 1) found a main effect of block, Greenhouse–Geisser corrected: $F(3.12, 71.67) = 3.34, p = .023, \eta^2_g = 0.06$, as well as a main effect of trial, $F(1, 23) = 49.56, p < .001, \eta^2_g = 0.07$, but no interaction between the two, $F(4, 92) = 0.79, p = .533, \eta^2_g = 0.01$.

**Trustworthiness ratings.** The changes in trustworthiness ratings for the faces in Experiment 4 are shown in Figure 3e. A repeated-measures ANOVA with rating time (pre- and postexperiment) and position of face in the task sequence (switch and repeat) as within-subjects factors found an effect of time with more positive trustworthiness ratings after the experiment than before, $F(1, 23) = 3.34, p = .07, \eta^2_g = 0.01$, but no significant effect of time in this experiment, $F(1, 23) = 0.69, p = .413, \eta^2_g = 0.01$, and no interaction between the two, $F(1, 23) = 0.18, p = .672, \eta^2_g = 0.00$.

To explore the contrasts between Experiments 1 and 5, we examined changes in ratings for valid faces and invalid faces separately. In this contrast, the assumption of normality was not violated, and so independent $t$ tests were used. When examining the change in trustworthiness ratings for low-fluency faces (switch and invalid faces), there was a significantly greater decrease in trust ratings during gaze cueing than task switching, $t(46) = -3.70, p = .001$, but no such difference emerged for valid faces, $t(46) = -0.50, p = .622$.

---

4 Comparing Experiments 1 and 5 in a 2 × 2 mixed ANOVA found a main effect of fluency (high—valid or repeat; low—invalid or switch), $F(1, 46) = 6.72, p = .013, \eta^2_g = 0.07$, as well as an effect of experiment, $F(1, 46) = 8.02, p = .007, \eta^2_g = 0.08$, and a significant interaction between the two, $F(1, 46) = 4.88, p = .032, \eta^2_g = 0.05$. Planned comparisons showed that high fluency (valid and repeat) faces did not differ significantly between the gaze-cueing and task-switching paradigms ($p = .509$), but there was a highly significant difference between low-fluency (invalid and switch) faces ($p < .001$). We did not include a similar analysis for Experiment 4 because it was much more methodologically distinct from Experiment 1 and so such comparison would be difficult to interpret.
Although the results of Experiment 5 showed a significant effect of face position, this appeared to be due to chance differences in the preratings—there is no logical reason to suppose that visuomotor fluency could have an effect before participants encountered it, and so these differences must have been due to random chance. The fact that they did not change over the course of the experiment, as evidenced by the lack of interaction and the remarkably flat profile of changes, is evidence that this incidental learning cannot be explained in terms of visuomotor fluency—even when accounting for whether the faces are targets or distractors.

However, the lack of any effect of time in this experiment suggested that the overall increase in trustworthiness seen in Experiment 4 may indeed have been due to the faces’ status as target rather than distractor stimuli, and reflect a familiarity effect not evident here.

**General Discussion**

Detecting and learning about subtle cues to trustworthiness is critical during social interactions. One such cue is the eye-gaze pattern of another individual, whether they are reliable and look toward relevant objects in a scene or deceive by looking away from objects. As shown in the RT results of the first three experiments, gaze cues are encoded rapidly and automatically and, hence, are effective ways of misdirecting others, because shifts of attention are difficult to inhibit. Our series of experiments further investigated the boundary conditions of the learning of trust from patterns of eye gaze.

Experiment 1 demonstrated that learning of trust is possible even when faces express neutral emotions. Previous work has highlighted the role of emotion in these gaze-trust effects. Bayliss et al. (2009) demonstrated significant trust learning effects when faces smiled, no effects when they frowned, and marginal effects when the faces were neutral. Experiment 1 revealed that significant trust effects can be obtained with neutral faces. It is unclear whether the failure to detect effects in the Bayliss et al. study was a Type I error or whether the changes to the procedure were of critical importance. In the previous work, a 2AFC task was employed where pairs of faces that had consistently looked toward
targets (valid) or had looked away from targets (invalid) were presented and participants selected the one who they felt was more trustworthy. In contrast, the current study required assessment of trust for each individual face and it measured changes in trust ratings from the start to the end of the experiment.

We believe this new approach is a more sensitive and robust means of measuring trust. Furthermore, it provides important information concerning where the effect might lie. That is, 2AFC can only reveal faces that previously looked toward targets tend to be selected as more trustworthy, not whether valid faces are trusted more, invalid faces trusted less, or both. The results of Experiment 1 suggest an asymmetry, where the effect is only observed in the decline in trust of invalid faces that looked away from targets, whereas there is no change in trust rating for the valid faces that always looked toward targets.

That this effect initially manifested as a decrease in trust only for invalid faces may relate to the finding by both Bayliss and Tipper (2006) and Bayliss et al. (2009) of a memory bias for invalid faces. That is, the false belief that invalid faces appeared more frequently than valid faces during the course of the gaze-cueing experiment. This builds on other work that has shown memory advantages for cheaters (Bell et al., 2012; Buchner, Bell, Mehl, & Musch, 2009), and has suggested that the motivation behind this trust effect is to remember those faces that frequently present challenges to visuomotor fluency—therefore, there is less change in trust for valid faces simply because participants are not motivated to remember those identities as clearly.

The second aim of this article was to explore the role that emotion plays at initial encoding. Experiment 2 demonstrated that there is a change in the pattern of trust learning when faces smile; that is, in contrast to Experiment 1, where effects were only detected in a decline in trust for invalid faces. When the faces smile, a bidirectional effect is observed where invalid faces again show a decrease in trust and valid faces now produce a significant increase in trust. This latter bidirectional effect with smiling faces has also been demonstrated by Manssuer and colleagues (Manssuer, Pawling, et al., 2015; Manssuer, Roberts, & Tipper, 2015).

There are many potential explanations for the difference in the pattern of results between Experiments 1 and 2 that future research should investigate. For example, the default learning mechanism might be to detect deception. Certainly in terms of memory for faces, this is better for faces that deceive (Bayliss et al., 2009; Bayliss & Tipper, 2006; Bell et al., 2012; Buchner et al., 2009), hence learning of trust is only evident in invalid faces that deceive and look away from targets. In contrast, when the faces all express positive emotion, this combines with the positive signal of joint attention evoked by valid cueing faces, hence increasing trust of these faces. Alternatively, the positive social context motivates participants to remember the faces in the experiment. Because invalid faces are apparently remembered well regardless of emotion (given the similar trust change profile for invalid faces across Experiments 1 and 2), this seems to primarily affect valid faces.

Our aim in the second half of this article was to explore some boundary conditions of this effect. Experiment 3 replaced the question of trustworthiness that participants were asked with a question of likability; simply by changing a single word in the design the effect was abolished. The lack of an effect when judging liking is somewhat counterintuitive. Therefore it was of value to report a further experiment in the supplemental materials. This study had a number of procedural differences to Experiment 3, the most important of which was that the faces expressed positive emotions (slight smiles). This situation is closer to Experiment 2 where positive smiles were observed, and again failed to show any effects when assessing liking of another person. Hence we doubt this was a Type II error.

This lack of effect with liking judgments suggests that the gaze-contingent trust effect is highly specific to trust—a fact that makes sense if one considers that trust as a trait judgment serves much more heavily as a predictive model of behavior than does liking: We decide how much to trust someone based on how we expect them to behave, whereas liking is a more subjective and affective judgment, and one less based in statistical contingencies. For example, incidental learning of gaze contingencies will influence economic decisions to invest in another person (e.g., Rogers et al., 2014). It is possible that if we manipulated participants’ beliefs about intentions or sense of competition that an effect of liking would emerge, but, at its basic level, this effect appears to be specific to monitoring the trustworthiness of interactants. To our knowledge, this provides the first evidence of a functional distinction between trust and liking, and directions for future research may examine other ways in which these two differ, and the possible mechanisms underlying each.

It is now well established that the eye movements of another person automatically shift attention, and whether they consistently look toward or away from objects mediates incidental learning of trust. The shifts of attention of another person certainly can be used to deceive, and hence it might be predicted that if there are no such behaviors in a face, then learning of trust does not take place, even though particular face identities are associated with different levels of visuomotor fluency. Therefore, Experiments 4 and 5 examined whether learning of trust could be generated in the absence of any physical changes to the faces through a task-switching procedure. We found that, in the absence of any physical changes, disruptions to participants’ sense of visuomotor fluency were not sufficient to generate changes in trustworthiness, despite the RT costs associated with task switching being comparable to those associated with gaze cueing (see Figure 2). This finding also held true whether or not the faces were targets (Experiment 4) or distractors (Experiment 5).

This contrasts with previous work that has shown perceptual fluency does increase liking of objects (e.g., Reber et al., 1998; Zajonc, 1968). It is also worth noting that this cannot be explained by the faces being more resilient to devaluation in Experiment 4, because Experiment 5 used the same faces as distractors that appeared before the target object. Taken together, these two results provide strong evidence against disruptions to visuomotor fluency being sufficient for incidental learning of trust.

There is also literature that has reported processing fluency can affect judgments of trust. For example, Winkielman, Olszanowski, and Gola (2015) found that increasing the disfluency associated with certain faces in an emotion categorization procedure led to decreased ratings of trust in later judgments, and that the effect of this disfluency was unrelated to face valence. However, it is important to note that our experiments examined learning in the absence of physical cues to trustworthiness (e.g., changes in expression) and, as such, these results are not necessarily inconsistent with this previous literature. Our interpretation fits with our earlier point that this incidentally learned trust reflects a sense of reliabil-


Received June 18, 2015
Revision received February 1, 2016
Accepted February 2, 2016