Cognition 108 (2008) 796–803

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Friend or foe: The effect of implicit trustworthiness judgments in social decision-making

M. van 't Wout *, A.G. Sanfey

Neural Decision Sciences Laboratory, Department of Psychology, University of Arizona, 1503 E University Boulevard, P.O. Box 210068, Tucson, AZ 85721, USA

ARTICLE INFO

Article history:
Received 30 June 2007
Revised 8 July 2008
Accepted 14 July 2008

Keywords:
Trustworthiness
Trust Game
Decision-making
Social
Emotional
Facial appearance
Automatic processing
Implicit processing

ABSTRACT

The human face appears to play a key role in signaling social intentions and usually people form reliable and strong impressions on the basis of someone's facial appearance. Therefore, facial signals could have a substantial influence on how people evaluate and behave towards another person in a social interaction, such as an interactive risky decision-making game. Indeed, there is a growing body of evidence that demonstrates that social behavior plays a crucial role in human decision-making. Although previous research has demonstrated that explicit social information about one's partner can influence decision-making behavior, such as knowledge about the partner's moral status, much less is known about how implicit facial social cues affect strategic decision-making. One particular social cue that may be especially important in assessing how to interact with a partner is facial trustworthiness, a rapid, implicit assessment of the likelihood that the partner will reciprocate a generous gesture. In this experiment, we tested the hypothesis that implicit processing of trustworthiness is related to the degree to which participants cooperate with previously unknown partners. Participants played a Trust Game with 79 hypothetical partners who were previously rated on subjective trustworthiness. In each game, participants made a decision about how much to trust their partner, as measured by how much money they invested with that partner, with no guarantee of return. As predicted, people invested more money in partners who were subjectively rated as more trustworthy, despite no objective relationship between these factors. Moreover, the relationship between the amount of money offered seemed to be stronger for trustworthy faces as compared to untrustworthy faces. Overall, these data indicate that the perceived trustworthiness is a strong and important social cue that influences decision-making.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Research that investigates social behavior and human interaction often focuses on the processing of facial features. Indeed, the face is an excellent source of information and in particular appears to play a key role in signaling social intentions from which people infer meaning, such as personality traits and complex social characteristics (Frith & Frith, 1999). Generally, people form reliable and strong impressions on the basis of facial appearance. In addition, facial cues related to threat are generally processed quickly and automatically, with only 40 ms exposure enough to generate an opinion about someone's intentions (Bar, Neta, & Linz, 2006). Moreover, there is ample evidence that the human brain has specialized areas for the processing of faces (Kanwisher, McDermott, & Chun, 1997). Therefore, facial signals could have a substantial influence on how people evaluate and behave towards another person in a social interaction, such as an interactive risky decision-making game.

A particularly useful task in examining interactive decision-making is the well-studied Trust Game (Berg, Dickhaut, & McCabe, 1995). In a standard Trust

* Corresponding author. Tel.: +1 520 6268597; fax: +1 520 6219306.
E-mail address: maschavantwout@gmail.com (M. van 't Wout).

0010-0277/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2008.07.002

Game two anonymous persons, an investor and partner, interact with one another. The investor is endowed with a certain amount of money and informed that she has the opportunity to transfer some amount of this endowment to her partner. She can transfer no money, the entire endowment, or any amount in between. The money transferred by the investor is then multiplied by the experimenter (usually by a factor of 3 or 4) and given to the partner. The partner, with whom the investor is not allowed to communicate during the game, then has the opportunity to transfer back some of the multiplied amount to the investor. In essence, the partner has a choice between honoring and abusing trust. If the partner honors trust and returns more money than transferred by the investor, both players end up with a higher monetary payoff than was originally obtained. However, if the partner abuses trust, the investor ends up with less money than the initial endowment, whereas the partner ends up with a large profit.

In a standard Trust Game, the investor and partner interact only once and therefore sharing the initial endowment is risky for the partner. From a traditional Game Theoretic standpoint, a rational (and selfish) partner would never honor trust given by the investor, and would keep the entire transferred amount for themselves. The investor, herself being rational, naturally realizes this, and so no amount would ever be sent over to the partner. Thus, by this theoretical view, trust is neither given nor repaid during this game (Berg et al., 1995; Camerer, 2003). However, in most experimental studies of the Trust Game, investors are usually willing to send some amount of money to the partner, and generally this trust is repaid to some extent. For instance in the original study by Berg et al. (1995), 30 out of 32 people did send some money to their partners, with on average half of the initial endowment being invested. In return, this trust was repaid with approximately one third of the multiplied amount, although there was a wide distribution with almost half of people sending back very little or nothing. Recently, a growing body of research acknowledges the role of social behavior in why people deviate from traditional Game Theoretic predictions (Camerer, 2003), and raises important questions as to exactly how social interaction can influence decision-making.

For instance, research has demonstrated that the willingness of the investor to trust can be altered by several experimental factors. Specifically, labeling the investor's partner with terms that imply cooperation or competition can affect decision-making, such as when the term 'opponent' is used instead of 'partner'. Use of the former resulted in half as much trusting behavior as the latter (Burnham, McCabe, & Smith, 2000). Information about the partner, such as how they treated someone else in the past (King-Casas et al., 2005), can also change the behavior of the investor. Studies have also examined how explicit information about the social and moral status of the partner can influence the investor's willingness to place trust (Delgado, Frank, & Phelps, 2005; Singer, Kiebel, Winston, Dolan, & Frith, 2004). Delgado et al. (2005) showed that investors were more likely to transfer greater amounts of money to partners described in vignettes as having high moral character as opposed to those demonstrating more neutral or more venal behavior. In this study, participants also gave higher trustworthiness ratings to the morally 'good' partners.

However, even in Trust Games where investors know nothing about the moral character or previous performance of their partner but are simply paired with an anonymous partner, there may well be substantial differences in terms of how much money is transferred to certain partners. Investors do make a (usually very rapid) decision regarding the trustworthiness of a previously unencountered partner and in the absence of any other information, it seems likely that automatically processed cues that signal one's social intentions, attitudes, or behavior could be vital in determining the likelihood of cooperation in the Trust Game. Since the face is particularly important in determining someone's intentions as outlined above, facial signals could have a substantial influence on people's behavior in the Trust Game. For instance, emotional expressions in general (Eckel & Wilson, 1999), and smiling in particular (Scharlemann, Eckel, Kacelnik, & Wilson, 2001), are social-emotional cues that could influence the decision to trust in economic games. Despite this, relatively little is known about how non-emotional, purely social, facial signals, such as competence or trustworthiness, influence strategic decision-making.

Trustworthiness in particular would appear to be a social facial signal of special significance, as it can potentially provide information about whether another individual is someone to approach or avoid, trust or distrust. Moreover, the trustworthiness of someone is readily judged from appearance with high reliability across people (Berry & Brownlow, 1989; Berry & McArthur, 1986; Brownlow, 1992). Indeed, Willis and Todorov (2006) have shown that people are able to judge the trustworthiness of faces very quickly (within 100 ms) and that this judgment is quite invariant even when more time is provided, suggesting that people are particularly efficient in judging the trustworthiness of faces. Though 'trustworthiness' is a rather ambiguous term, it has been argued that the detection of trustworthiness is essential for human survival (Cosmides & Toobey, 2000), and there is good evidence for a set of brain regions which may be involved in the perception of trustworthiness. An fMRI study by Winston, Strange, and O'Doherty Dolan (2002) showed that facial trustworthiness evaluation is associated with activation of the amygdala and insula, specifically that these areas are more engaged as the untrustworthiness of a face increases. This is particularly interesting, as these brain areas have been implicated in detecting potentially dangerous stimuli (Amaral, 2002), encoding feelings of disgust (Phillips et al., 2004) and in other studies of (social) strategic decision-making (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003).

The relationship between these highly automatic, rapid, judgment processes and strategic decision-making is clearly relevant in many real-world interactions. Though it seems likely that perception of an individuals' trustworthiness is related to decisions made about that person in strategic interactions, this question has surprisingly not

been studied in the laboratory to date. The present study sought to remedy this, and examined the role of perceived trustworthiness on risky social decision-making. Participants played the role of investor in a one-shot, two-person Trust Game with partners who had been previously independently rated on trustworthiness. We hypothesized that investors who are confronted with partners perceived as less trustworthy would invest less money than when confronted with more trustworthy-looking partners. After the Trust Game rounds participants saw their partners again and rated their trustworthiness. In addition, we were interested in whether the relationship between judged trustworthiness and trust behavior might be is stronger for faces that are perceived as either trustworthy or untrustworthy. This question is motivated by the fMRI study of Winston et al. (2002) who reported amygdala activation particularly in response to untrustworthy faces, suggesting that untrustworthiness might be a stronger facial cue than trustworthiness. On the other hand, in an fMRI study by Singer et al. (2004) participants showed brain activation in areas associated with social cognition particularly in responses for partners that had shown to be cooperative in a Prisoner's Dilemma Game, which is similar in structure to the Trust Game. Moreover, this study design also allowed a preliminary analysis whether people recognize particular faces better than others. More specifically, Singer et al. (2004) showed that memory performance was better for faces that were associated with cooperation than with defection. Thus, we hypothesized that recognition would be better for faces of partners that repaid trust as compared to those that abused trust, and additionally that this might be mediated by perceived trustworthiness, i.e., recognition memory is best for subjectively trustworthy faces that also acted trustworthy.

2. Methods

2.1. Participants

Sixty undergraduate students (33 women, 27 men) participated in this study. Two individuals were excluded from the analyses as they did not follow the rules of game properly, i.e., offering more to the partner than was allowed or providing no trustworthiness ratings. This resulted in a group of 32 women and 26 men. The number of men and women that participated in the experiment did not differ significantly from each other, $Chi^2 = 0.62$, $p = .43$. Mean age was 18.7 years ($SD = 1.4$). The study was conducted in compliance of the Declaration of Helsinki and local ethics committee approval, and all participants provided written informed consent.

2.2. Trust Game

Participants were instructed as to the nature and rules of the Trust Game. Participants acted only in the role of investor in the Trust Game. In the task instructions it was emphasized that the participant's partners in the game would play the game independently of each other, with no collusion. Participants were also told that in addition to course credit they would be paid based on their choices in the game. After finishing the game, they received, in cash, a small percentage of their overall game earnings (average of $4).

Participants played against 79 different partners, and began each trial by being endowed with $10. On each trial, participants could decide to transfer over any amount of this money, in one dollar increments. That is, they decided to send over some amount between $0 (placing no trust) and $10 (placing full trust). If they decided to trust the partner, the amount of money selected was multiplied by four and transferred to the partner.

Partners responded in pre-programmed fashion. When a non-zero amount was transferred by the investor (that is, some trust was placed), 50% of the time this trust was repaid and 50% of the time this trust was abused. When trust was repaid, the partner sent back half of the transferred, multiplied, amount to the investor. For example, if $4 was transferred initially, this was multiplied by 4 for a total of $16, and then half of this (i.e., $8) was sent back. The investor therefore earned this $8, plus her original $6 (the $10 endowed minus the transferred $4), for a total of $14 on this trial.

When trust was abused the exact outcome depended on the amount of money transferred to the partner. When the participant transferred less than $5, nothing was returned back. If the transferred amount was more than $5, half the amount money the participant initially invested (12.5% of the total multiplied amount) was sent back. In both of these cases, the investor ultimately earned less than her originally endowed $10.

Before each investment period participants were presented with a picture of their partner for that trial. Pictures were 79 grayscale frontal photographic images of emotionally neutral faces that were selected from a larger set of 150 images; 100 are from the set of Adolphs, Tranel, and Damasio (1998), supplemented with 50 images from the psychological image collection of the psychology department of Stirling University (http://pics.psych.stir.ac.uk/). Faces were selected on the basis of trustworthiness ratings given by 36 healthy subjects in a separate pilot study (9 men, 27 women; mean age 21.6, $SD = 3.3$). All faces were randomly presented in the game.

Following presentation of the photograph, participants then decided how much of their endowment to offer to their partner on that trial and pressed the appropriate key on the computer keyboard to send the money over. After a short delay, the participant was informed of their partner's decision and the total amount earned in the trial was shown (see Fig. 1 for a trial timeline).

2.3. Subjective trust rating and recognition memory

After playing the Trust Game, participants completed an unrelated task for approximately 20 min. Following that, participants were presented serially with 150 faces, comprised of the 79 faces that were presented in the Trust Game and 71 decoy faces. All 150 faces were randomly presented. They were asked to rate how trustworthy they thought each face was on a scale of "1" (highly untrustworthy) to "7" (highly trustworthy), with "4" being neutral, and also to rate whether they had seen that face before,

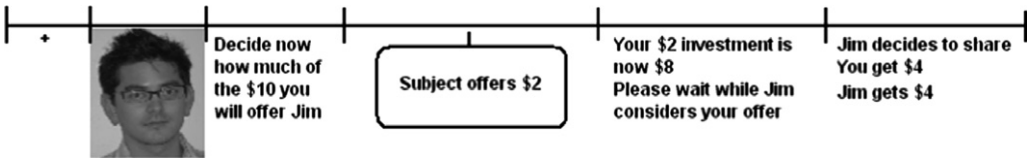

Fig. 1. Single trial in the Trust Game.

on a scale from "1" (definitely not seen) to "5" (definitely seen). Due to technical problems, six people failed to complete this part of the experiment.

3. Results

3.1. Trust behavior and trustworthiness

Overall, participants placed some amount of trust in 85% of trials. Nine participants placed some trust in every partner, i.e., sent over some amount of the $10 on every round, whereas the remaining 49 participants sometimes sent nothing to their partner. Although there was variation in what was sent to a partner by different participants, most participants used the full range from $0 to $10, and only nine participants never sent more than $5 (half their endowment) to their partners. The mean amount of money offered when trust was placed was $3.52 (*SD* 2.65, range $0-$10). On average, less money was sent over to a partner at the end of the game (last 20 rounds) as compared to the beginning of the game first (20 rounds), $t(57) = 3.39$, $p = .001$. Mean subjective trustworthiness rating of faces of partners after the game was 4.1 ($SD = 1.65$, range 1–7). Subjective trust ratings given by participants correlated extremely highly with trust ratings of the norm group, Spearman $r = .89$, $p < .0001$.

There was a significant correlation between the normed trustworthiness rating of the partner and the amount of money that was transferred by the participant, Spearman $r = .75$, $p < .0001$ (see Fig. 2). In addition, there was a significant correlation between subjective trustworthiness rating of the partner and the amount of money transferred by the participant, Spearman $r = .82$, $p < .0001$. This result strongly suggests that participants were indeed influenced by the perceived subjective trustworthiness of the partner in their judgment of whether to trust, and how much trust to place. Of additional interest was whether the amount transferred to a partner is also influenced by the outcome experience by the participant (trust repaid or abused) in the previous round. Therefore we performed a mixed linear model with subjects as random factors to test whether the amount of money sent over to a partner was influenced by normed trustworthiness ratings, the outcome on the previous round (trust repaid or abused) and their interaction. This resulted in a significant effect of normed trustworthiness ratings, $F(1,4702) = 81.15$, $p < .0001$; outcome, $F(2,4702) = 6.66$, $p = .001$); but no significant interaction between normed trustworthiness ratings and outcome, $F(2,4702) = 0.61$, $p = .55$.

It is possible that other factors, besides trustworthiness, may play a role in the decision as to how much money to send over to the partner. Although we tried to eliminate possible factors such as facial emotional expression, which are known to influence decision-making, participants could have picked up other facial features, such as attractiveness. Therefore we collected attractiveness ratings of the set of faces, rated by an independent sample of undergraduate students who did not play the game. Results showed that although attractiveness correlated with trustworthiness, Spearman $r = .55$, $p = .0002$, a linear mixed model analysis showed that attractiveness was not correlated with the amount of money sent over to that partner, $F(1,2266) = 0.26$, $p = .61$, whereas trustworthiness ratings were $F(1,2266) = 9.78$, $p = .002$.

In their fMRI study, Winston et al. (2002) reported that untrustworthy faces in particular appeared to activate emotional brain areas, suggesting that the perceptual system may be especially sensitive to untrustworthy faces. To explore this question in the context of decisions made in the Trust Game, we split the group of faces on the basis of scores given by participants into untrustworthy (score below neutral, "4") and trustworthy (score above neutral). Interestingly, the correlation between subjective trustworthiness and amount of money transferred was stronger for partners rated in general as trustworthy, Spearman $r = .65$, $p < .0001$ ($n = 45$), as opposed to those rated as untrustworthy, Spearman $r = 0.29$, $p = .10$ ($n = 34$). To test whether these two correlations differ significantly from one another we performed a Fischer z transformation for independent samples. Using the participant trust ratings, the correlations were significantly different ($p < .05$). Similar results were found when we used normed trustworthiness ratings (correlation with amount of money transferred was Spearman $r = 0.45$, $p = .001$ and Spearman $r = 0.19$, $p = .29$ for trustworthy and untrustworthy faces, respectively) and this difference between the correlation reached trend level significant ($p < .08$). It is important to note here that the number of faces rated as untrustworthy is smaller than the number of faces rated as trustworthy. However, the range of scores for trustworthy and untrustworthy faces is comparable, i.e., 1.20 (variance = 0.092) and 1.31 (variance = 0.086), respectively. This suggests that the lack of correlation between untrustworthy faces and amount of money transferred is probably not due to a smaller sample size or a restricted range in ratings.

Because we measured subjective trustworthiness after playing the Trust Game, playing the Trust Game itself could influence subjective trustworthiness ratings. For instance, merely seeing a given face twice (once in the game and once in the ratings test) may increase subjective trustworthiness ratings due to familiarity. Indeed, subjective trust ratings were significantly higher for faces that

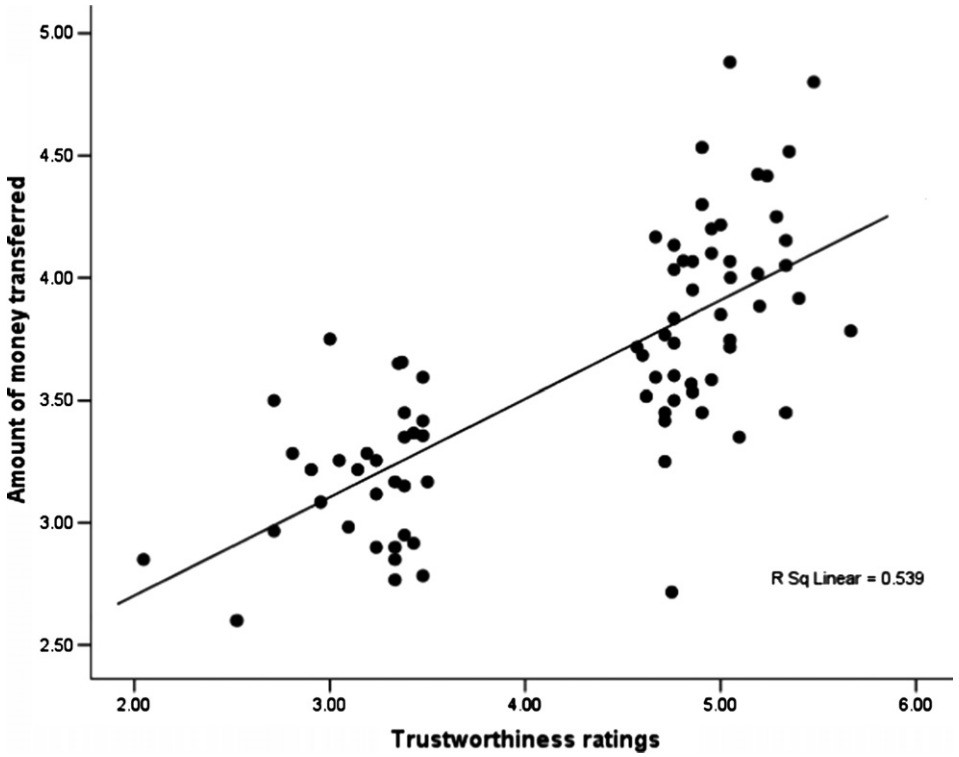

Fig. 2. Scatter plot for normed trustworthiness ratings of a particular face (horizontal axis) and the average offer amount send to that partner (vertical axis).

were seen previously than for the new decoy faces, $t(51) = -9.67$, $p < .0001$). Additionally, partners in whom trust was placed, but who did not reciprocate this trust, might be rated as less trustworthy afterwards. Although, subjective trustworthiness ratings were not significantly different for partners that repaid trust, as compared to abused trust, $t(57) = 0.12$, $p = .91$, we performed a linear mixed model analysis in which participants' trustworthiness ratings were regressed on normed trustworthiness ratings, the partner's return (calculated as the percentage of the amount of money sent back as a function of what was sent over), and their interaction, to specifically test whether experience with a partner influences subjective trustworthiness above and beyond normed trustworthiness ratings. This resulted in a significant relationship only between participant and normed trust ratings, $F(1,4132) = 448.05$, $p < .0001$. Neither the relationship between partner return, $F(3,4132) = 0.96$, $p = .41$, or the interaction between normed trust ratings and partner return were significant, $F(3,4132) = 0.43$, $p = .73$. This suggests that post-game participant trust ratings were specifically predicted by normed trust ratings.

Finally, there may be a general effect whereby trust increases or decreases across the whole experiment, independent of the actual faces and partner behavior. We did not find evidence of this, with no significant difference observed in subjective trust ratings for the first 20 partners as compared to last 20 partners faced in the Trust Game, i.e., trust ratings for faces at the end of the game are not higher than at the beginning, $t(52) = -0.98$, $p = .33$). Instead, only partners in whom no trust was placed were rated afterwards as significantly less trustworthy, $t(48) = -4.60$, $p < .0001$).

3.2. Recognition memory

With respect to memory for faces, faces that were presented in the Trust Game were given higher recognition ratings than faces that were new, $t(51) = 10.76$, $p < .0001$). As mentioned previously, we hypothesized that recognition would be better for faces that repaid trust compared to those that abused trust, that subjectively trustworthy faces would be recognized better than untrustworthy faces, and a potential interaction, namely that faces rated as trustworthy that subsequently repaid trust would have the highest recognition scores.

Recognition was indeed better for faces that were rated as trustworthy by participants as compared to faces that were rated as untrustworthy, $F(1,50) = 35.59$, $p < .0001$). However, recognition ratings for faces that abused trust did not differ significantly as compared to faces that repaid trust, $F(1,50) = 0.01$, $p = .92$), nor was there a significant interaction between trustworthiness of the face and decision of the partner, $F(1,50) = 1.14$, $p = .29$) (see Fig. 3). This suggests that a partner who is a priori thought to be trustworthy is remembered better than a partner who is a priori thought to be untrustworthy, regardless of their actual behavior in the game. However, it is possible that subjectively trustworthy faces are perceived as more familiar because they are perhaps more average or attractive

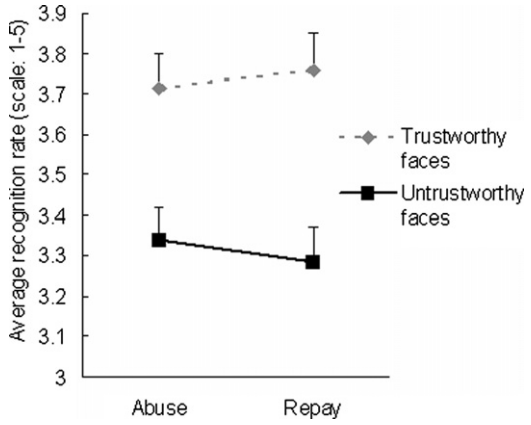

Fig. 3. Recognition ratings (SE) for faces when trust was abused or repaid for trustworthy and untrustworthy faces.

compared to the untrustworthy faces. Therefore, we used signal detection methods to calculate the false alarm rate (saying a face is remembered whether it was absent in the task), the hit rate (correct recognition of a face), false negative rate (saying a face is new when it was present in the task), correct rejection, and d' (sensitivity of recognition memory). Results showed that there is no difference between d' for trustworthy (0.76) and untrustworthy faces (0.79), $t(49) = 0.49$, $p = .62$, suggesting there is no difference between sets of faces in recognition sensitivity. However, it should be noted that the overall the recognition sensitivity is quite low with a hit rate of 67%, a correct recognition rate of 60%, a false negative rate of 33% and a false alarm rate of 39% across all faces.

4. Discussion

In this study individuals played a standard, one-shot, Trust Game with 79 hypothetical partners, each of whom had been previously rated by an independent group on their subjective trustworthiness. Results demonstrated that implicit processing of social cues, i.e., trustworthiness, had a reliable effect on decisions made in the Trust Game, namely that participants chose to invest more money with those partners who had higher trustworthiness ratings. These results suggest that an initial impression of trustworthiness is a prevailing cue that is automatically, and rapidly, processed (Willis & Todorov, 2006; Winston et al., 2002), and can extend a strong influence on decision-making. Interestingly, despite playing the game with each partner and thereby obtaining direct evidence of the trustworthiness of the person, the post-game trustworthiness ratings of the partners were not influenced by whether the partners had in fact abused or repaid trust. Instead, it appears that the rapid, intuitive judgment of trustworthiness based on facial appearance exerts a rather dominant influence. Moreover, it is important to realize that the amount of money that was transferred to a partner correlated highly with both ratings of trustworthiness given by an independent sample of subjects who did not play the game, as well as the subjects that did play the game. Despite the trustworthiness ratings of partners, the outcome on a previous round (whether trust was repaid or abused) also had an effect of how much money as send over. Thus after trust was abused people send over less money in the next round. This is in line with our finding that participants send over less money at the end of the game compared to the beginning of the game. This thus may have biased participants towards making less use of the facial features of trustworthiness (and untrustworthiness) on each trial. Yet, we observed a strong significant influence in the amount of money transferred to partners based on their facial features of trustworthiness. Additionally, the effect of outcome on the amount of money send over did not interact with trustworthiness ratings, which suggests that although participants changed their strategy throughout the game, the effect of perceived trustworthiness was robust.

The robustness of this influence is further demonstrated by the strong correlation between normed trustworthiness ratings (from an independent sample of raters) and trustworthiness ratings obtained from our participants. The strong relationship between subjective trustworthiness and amount of money transferred to a partner is in accordance with recent studies that show that the moral and social status of a hypothetical partner can influence behavior in economic decision-making using similar tasks (Delgado et al., 2005; Singer et al., 2004). However, our findings extend these studies in demonstrating that even non-explicit, automatically processed, social information signaled by facial features appears to be a strong cue in influencing decision-making in an interactive social context. In addition, our results are in line with, and indeed extend, previous research showing that facial cues, such as smiling, influences strategic decision-making (Eckel & Wilson, 1999; Scharlemann et al., 2001). These results could have real-life implications to what extent facial features guide us in real-life decision-making, such as negotiation, and how much control we have over these influences. A previous study investigating the strength of character traits derived from facial features showed that although impressions are particularly strong other sources of information could change these impressions (Hassin & Trope, 2000).

Since it has been demonstrated that strong and reliable impressions, such as trustworthiness, are derived from the face (Bond, Berry, & Omar, 1994; Hassin & Trope, 2000; Willis & Todorov, 2006), it is interesting to know whether people are more sensitive to positive (trustworthiness) or negative (untrustworthiness) facial cues. Here, the influence of subjective trustworthiness on economic decision-making appeared to be strongest in the most trustworthy faces. This finding was not due to a restricted range in ratings of untrustworthy faces. These results are in line with the interpretation given by Singer et al. (2004) who found that cooperators in particular elicited more activation in social-emotional brain areas, such as the amygdala and insula, suggesting that trustworthiness might be a more important social cue than untrustworthiness. Nevertheless, the converse, i.e., a stronger correlation between investment and untrustworthy faces, could also been hypothesized. For instance, Winston et al. (2002) found more activation of social brain areas when people judged

subjectively untrustworthy faces as compared to trustworthy faces. An interpretation of this finding is that the human brain reacts in particular to untrustworthiness, as it is a more salient threat-related social cue that might therefore influence decision-making in a social interaction to a stronger degree than trustworthiness. This leads to the question of whether people are more sensitive to positive or negative feedback, a question also addressed by both Singer et al. (2004) and Rilling et al. (2002). Some proposed models of facial processing in humans suggest that this skill evolved because it allows individuals to keep track of others who had proved to be cooperative in interactions (Cosmides & Toobey, 2000). According to this account, trustworthy partners in particular should be remembered better and offered the most money.

With respect to memory, our data shows that although recognition memory ratings for subjectively trustworthy partners were higher than for subjectively untrustworthy partners, the signal detection analysis shows that there was no difference between participants' ability in recognizing trustworthy and untrustworthy faces. However, it must be noted that the hit rate and correct rejection rate of all faces was 67% and 60%, respectively, which shows that participants were relatively poor at recognizing their partners. In addition, we found no influence of experience in the game on the ability to remember partners. This is contrary to our hypothesis that participants would recognize partners who repaid trust better than those who abused trust (based on Singer et al., 2004) who showed that memory for cooperative partners was better than for defectors. The present data are also not in line with Mealey, Daood, and Krage (1996) or Yamagishi, Tanida, Mashima, Shimoma, and Kanazawa (2003), who reported that people who showed to be untrustworthy (by cheating or not cooperating) are remembered better, even when they do not know who are cheaters and cooperators. On the other hand, in a series of well controlled experiments, Mehl and Buchner, (2008) recently reported that there was no difference in recognition memory between faces associated with a history of cheating and faces associated with a history of trustworthiness. Conversely, our results suggest that the perception of subjective trustworthiness may in fact be stronger than the actual game experience. However, interpretation of these conclusions should be made with caution, as the present study was not designed to exclusively focus on memory. For example, the large amount of faces the participants were exposed to ($n = 79$) might have made it too difficult to keep track of the experience with each partner, and therefore may have diluted the effect of experience on recognition memory. Indeed, as mentioned earlier we found low correct recognition rates across all faces. In addition, we did not ask participants specifically whether they remembered if each partner either abused or repaid their trust. Taken together, we believe it is still unclear in what situation people are more likely to attend to defectors or to cooperators. Future research should investigate this interesting question in more detail. For instance, these results most likely change drastically when people play a repeated-round game with partners as in such a situation memory for previous outcomes, i.e., reputation is important, whereas in the present single-shot version we highlight the fact that after a decision is made by a partner the game is over. For these reasons future research is needed to investigate the interaction between first impressions based on rapid visual presentation and the actual outcome experienced with that partner on memory and behavior in (sequential) strategic decision-making. Further, the issue of whether cooperators or defectors are better encoded and remembered is likely to be related to personality factors (some people learn better from negative feedback and others from positive feedback) (Carver & White, 1994; Zuckerman, Joireman, Kraft, & Kuhlman, 1999) and perhaps even the specific task used.

It is worth noting that although we report a strong relationship between subjective trustworthiness and trust behavior in the Trust Game, we did not specifically investigate other social facial cues that could affect behavior in the Trust Game. For instance, Willis and Todorov, (2006) showed that besides trustworthiness, facial features such as attractiveness, liking, competence, and aggressiveness are processed fast and automatically. Although the present set of faces was controlled for emotional expression, the faces were initially not rated on these other facial features. However, since previous studies showed that more attractive people are more trusted (Mulford, Orbell, Shatto, & Stockard, 1998; Solnick & Schweitzer, 1999; Wilson & Eckel, 2006), we asked an independent group of undergraduate students rate the faces on attractiveness. Although attractiveness correlated with trustworthiness, the attractiveness ratings did not correlate with the amount of money send over to that partner. Though we cannot rule out that other factors could have influenced decisions in this game, it seems plausible that many of these additional features could influence decision-making via the perception of trustworthiness. Future studies could investigate the influence of these automatically processed social signals in decision-making. Additional questions that need investigation refer to whether people are actually accurate in their assessment of trustworthiness, i.e., are partners that are perceived as more trustworthy actually in fact more trustworthy, are there certain markers for trustworthiness, such as the distance between particular facial features (e.g., eyes, eyebrows, mouth), or can trustworthiness be mimicked? Although it is not within the scope of present study to tackle the question of whether trustworthiness is a useful and accurate signal, studies that address the question of whether facial signals can be correctly mimicked showed that people are rather accurate in distinguishing a fake smile from a genuine one (Ekman & Friesen, 1982). Moreover, in a study on authentic smiling and customer satisfaction, results showed both higher perceived friendliness and increased customer satisfaction when servers' smiles were genuine (Grandey et al.,2005). Future studies should elucidate the accuracy of one's judgments as people might treat trustworthy-looking people consistently different, which can elicit the expected behavior of being more trustworthy.

In summary, we observed a strong relationship between the perceived trustworthiness of a partner and the amount of money that was transferred to this partner by our participants. These data suggest that even when people

are not explicitly asked about subjective facial trustworthiness of a partner they seem to evaluate this social cue implicitly and act upon it. This underlines the fact that automatically processed facial features of partners influence strategic decision-making, and that these processes should be integrated into models of cooperation and reciprocal behavior in interactive decision-making.

Acknowledgments

The authors thank Professor Ralph Adolphs for his generous permission to use the set of faces used by his research group and Luke Chang for his helpful comments. This study was supported by funding from the Netherlands Organisation for Scientific Research (NWO), Rubicon Grant (No. 446.05.003).

References

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature, 393*, 470–474.

Amaral, D. G. (2002). The primate amygdala and the neurobiology of social behavior: Implications for understanding social anxiety. *Biological Psychiatry, 51*, 11–17.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*, 269–278.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behaviour, 10*, 122–142.

Berry, D. S., & Brownlow, S. (1989). Were the physiognomists right? Personality correlates of facial babyishness. *Personality and Social Psychology Bulletin, 15*, 266–279.

Berry, D. S., & McArthur, L. Z. (1986). Perceiving character in faces: The impact of age-related craniofacial changes on social perception. *Psychological Bulletin, 100*, 3–18.

Bond, C. F., Berry, D. S., & Omar, A. (1994). The kernel of truth in judgments of deceptiveness. *Basic and Applied Social Psychology, 15*, 523–534.

Brownlow, S. (1992). Seeing is believing: Facial appearance, credibility, and attitude change. *Journal of Nonverbal Behavior, 16*, 101–115.

Burnham, T. C., McCabe, K., & Smith, V. L. (2000). Friend-or-foe intentionality priming in an extensive form trust game. *Journal of Economic Behavior & Organization, 43*, 57–73.

Camerer, C. F. (2003). Dictator, ultimatum, and trust games. In *Behavioral game theory* (pp. 43–116). New York: Russell Saga Foundation.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67*, 319–333.

Cosmides, L., & Toobey, J. (2000). The cognitive neuroscience of social reasoning. In M. S. Gazzaniga (Ed.), *The new cognitive neuroscience* (pp. 1259–1276). Cambridge; Massachusetts: MIT Press.

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience, 8*, 1611–1618.

Eckel, C. C., & Wilson, R. K. (1999). *Reciprocal fairness and social signaling: Experiments with limited reputations*. New York, NY: American Economic Association.

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior, 6*, 238–252.

Frith, C. D., & Frith, U. (1999). Interacting minds – A biological basis. *Science, 286*, 1692–1695.

Grandey, A. A., Fisk, G. M., Mattila, A. S., Jansen, K. J., & Sideman, L. A. (2005). Is "service with a smile" enough? Authenticity of positive displays during service encounters. *Organizational Behavior and Human Decision Processes, 96*, 38–55.

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology, 78*, 837–852.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17*, 4302–4311.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science, 308*, 78–83.

Mealey, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology, 17*, 119–128.

Mehl, B., & Buchner, B. (2008). No enhanced memory for faces of cheaters. *Evolution and Human Behavior, 29*, 35–41.

Mulford, M., Orbell, J., Shatto, C., & Stockard, J. (1998). Physical attractiveness, opportunity, and success in everyday exchange. *The American Journal of Sociology, 103*, 1565–1592.

Phillips, M. L., Williams, L. M., Heining, M., Herba, C. M., Russell, T., Andrew, C., et al. (2004). Differential neural responses to overt and covert presentations of facial expressions of fear and disgust. *NeuroImage, 21*, 1484–1496.

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron, 35*, 395–405.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 300*, 1755–1758.

Scharlemann, J. P. W., Eckel, C. C., Kacelnik, A., & Wilson, R. K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology, 22*, 617–640.

Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron, 41*, 653–662.

Solnick, S. J., & Schweitzer, M. E. (1999). The influence of physical attractiveness and gender on ultimatum game decisions. *Organizational Behavior and Human Decision Processes, 79*, 199–215.

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

Wilson, R. K., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly, 59*, 189–202.

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

Yamagishi, T., Tanida, S., Mashima, R., Shimoma, E., & Kanazawa, S. (2003). You can judge a book by its cover: Evidence that cheaters may look different from cooperators. *Evolution and Human Behavior, 24*, 290–301.

Zuckerman, M., Joireman, J., Kraft, M., & Kuhlman, M. D. (1999). Where do motivational and emotional traits fit within three factor models of personality? *Personality and Individual Differences, 26*, 487–504.