Developing Trust: First Impressions and Experience

Michael Yu[1], Muniba Saleem[2], and Cleotilde Gonzalez[1]

[1] Dynamic Decision Making Laboratory

Carnegie Mellon University

4609 Winthrop Street

Pittsburgh, PA 15213, USA

[2] University of Michigan, Research Center for Group Dynamics, Institute for Social Research and

University of Michigan-Dearborn, Department of Behavioral Sciences

4901 Evergreen Road

Dearborn, Michigan 48128, USA

Abstract

Using the repeated Trust Game, we investigated how first impressions and experience affect trusting dispositions, beliefs, and behaviors. As in previous research, trusting beliefs and trust-related behaviors were greater at the start of the game for partners with trustworthy faces; and higher later in the game for partners who reciprocated. Three additional findings extended beyond the previous research. First, by measuring the discrete components of trusting beliefs rather than an umbrella "trustworthiness" measure, we confirmed that first impressions and experience influence judgments of competence, benevolence, and integrity. Moreover, we found suggestive evidence that perceptions of benevolence and integrity updated more quickly with experience than perceptions of competence. Second, by looking at trusting beliefs at the start of two consecutive repeated Trust Games, we found that judgments of competence, benevolence, and integrity continue to be influenced by trustworthy facial appearances, even after previous beliefs based on facial appearances were disconfirmed. Third, we found increased investment with a partner at the start of a second repeated Trust Game, even when participants expected their partners to betray them. Overall, our results clarify our understanding of how first impressions and experience influence trusting beliefs; provides evidence that changes in the repeated Trust Game represents learning about a specific partner rather than revisions of trusting dispositions; and highlights important distinctions between trusting beliefs and trust-related behaviors.

1. Introduction

"You can't judge a book by its cover." This famous figure of speech has often been applied to interpersonal contexts, warning individuals not to judge others by simply relying on their external appearances. On the other hand, people frequently rely on facial appearance to draw trait inferences about others and subsequently use these judgments to guide their own behavior. For example, extensive reviews of research on physical attractiveness reveal that people attribute positive characteristics such as intelligence, competence, leadership skills, and trustworthiness to attractive persons (Eagly, Ashmore, Makhijani, & Longo, 1991; Feingold, 1992). Beyond influencing positive beliefs, physical attractiveness also results in obtaining better outcomes in most domains of life (Hamermesh & Biddle, 1994; Zebrowitz, 1999). Facial appearance has also been used to judge trustworthiness in others (van 't Wout & Sanfey, 2008). Interestingly, people are able to judge the trustworthiness of faces very quickly (within 100 ms) and this judgment is robust even when more time is provided (Willis & Todorov, 2006). In addition to influencing beliefs, facial appearance has been shown to be predictive of trusting behaviors as assessed by the Trust Game (Campellone & Kring, 2013; Chang, Doll, van 't Wout, Frank, & Sanfey, 2010; DeBruine, 2002; Eckel & Wilson, 2003; Scharlemann, Eckel, Kacelnik, & Wilson, 2001; van 't Wout & Sanfey, 2008). Overall, this work suggests that people often rely on facial appearance to assess the trustworthiness of their opponent and use their subjective perceptions to guide their decisions regarding whether or not to invest with this opponent.

However, facial appearance is not the only predictor of trust. Previous research has found that facial appearances often determine *initial* judgments of trustworthiness and trust-related behavior, while later judgments and behaviors are dictated by the participant's experiences with a specific trust partner (Campellone & Kring, 2013; Chang, Doll, van 't Wout,

Frank, & Sanfey, 2010). As such, trust is initially influenced by first impressions – snap judgments made based upon facial appearances – and subsequently determined by the interactive experience with a particular partner.

Unfortunately, these previous studies often measured trust using either an umbrella measure of "trustworthiness" or using trust-related behaviors. Such high-level measures provide only a partial understanding of trust (Ben-Ner & Halldorsson, 2010). In contrast, more precise models of trust that view judgments of competence, benevolence, and integrity as discrete components of trusting beliefs have been developed (Mayer, Davis, & Schoorman, 1995; McKnight & Chervany, 2001). Moreover, these components have been shown to differ in their causes and effects. For example, perceptions of competence were shown to influence the acceptance of tacit knowledge, i.e., knowledge that is difficult to articulate (Levin & Cross, 2004), and research on trust repair suggests that different interventions are required to address perceptions of low benevolence and integrity compared to perceptions of low competence (Kim, Dirks, Cooper, & Ferrin, 2006; Xie & Peng, 2009). Trust-related behaviors also depend not only upon characteristics of the person being trusted but the person offering their trust. Individuals who might offer their trust may differ in the degree to which they believe the competence, benevolence, and integrity of people *in general*. They may also use strategic or rule-based considerations to guide whether or not they want to offer their trust (McKnight & Chervany, 2001). Thus, while previous research suggests that facial appearance and experience can influence trust, the exact nature of their influence remains unclear.

In this paper, we investigate how first impressions and experience affect trust in a repeated Trust Game. First impressions refer to the snap judgments made regarding a person's trustworthiness based upon their facial appearance. Experience refers to repeated interactions

with a partner, including feedback on whether the partner tends to reciprocate or betray trust.

Trust is evaluated along three dimensions, including [1] *trusting beliefs*, i.e., perceptions of the

competence, benevolence, and integrity of a specific partner; [2] *trust-related behaviors*, i.e.,

actions that make oneself more vulnerable to others for a potential benefit; and [3] *trusting*

dispositions, i.e., attributes of the person engaging in trust that influence perceptions of the

competence, benevolence, and integrity of others in general (faith in humanity) or strategic and

rule-based decisions to engage in trust-related behaviors (trusting stance). This specification of

trust is adapted from a model developed by McKnight and Chervany (2001, 2002), which

consolidated different interpretations of trust that had been used across psychology, economics,

sociology, political science, management, and communications. In contrast to previous research,

our paper focuses on changes in the perceptions of competence, benevolence, and integrity

(Study 1 and Study 2) rather than an umbrella measure of "trustworthiness." In considering

trusting dispositions, we move beyond the previous research by asking whether experience

influences only trusting beliefs in individual partners or influences general dispositions towards

others (Study 2).

1.1 Trust model

McKnight and Chervany's trust model was developed to clarify and connect different dimensions

of trust that had been investigated in earlier research (McKnight & Chervany, 2001). In our paper, we

focus on a subset of their model that influences interpersonal trust – trusting dispositions, trusting beliefs,

and trust-related behaviors. Trusting dispositions reflect a person's general approach towards trust across

multiple contexts. Trusting beliefs reflect a person's perceptions of the trustworthiness of a specific

individual. Trust-related behaviors reflect actions that a person may take to obtain a potential benefit by

becoming vulnerable to another person. Trusting dispositions influence trust-related behaviors both directly and through changes in trusting beliefs. This abbreviated model is provided in Figure 1.

As noted, trusting dispositions reflect a person's trait-like tendencies towards trusting others. Trusting dispositions are comprised of two dimensions: faith in humanity and trusting stance. *Faith in humanity* reflects a person's general belief about another person's competence (a person's ability to achieve her goals), benevolence (the degree to which a person cares about others), and integrity (a person's adherence to prescriptive norms). If a person generally believes that others are high in competence, high in benevolence, and high in integrity, they will use these beliefs to guide their initial interactions with new people. *Trusting stance* reflects general strategies or principles that may guide a person's trust-related behavior. For example, a person may engage in trust-related behaviors strategically to gather information about another person's trustworthiness. Alternately, a person may simply adhere to philosophies that proscribe trust-related behavior, e.g., "do unto others" or "everyone deserves a second chance." Both faith in humanity and trusting stance are likely to determine initial trust-related behavior with strangers. However, faith in humanity produces trust-related behavior based upon the expectation that trust will be reciprocated, whereas trusting stance produces trust-related behavior based upon rules that do not require expectations of reciprocity. Trusting dispositions are represented on the left of the diagram in Figure 1.

Trusting beliefs reflect our perceptions of the competence, benevolence, and integrity of a specific person. These are the same dimensions underlying faith in humanity, but reflect beliefs about a specific individual rather than beliefs about others in general. Initial trusting beliefs are believed to derive from faith in humanity. However, trusting beliefs in a specific person are expected to diverge from faith in humanity after repeated experience. Previous research provides

evidence consistent with this update. For example, positive interactions in the Trust Game have led participants to report that their partner was highly "trustworthy" (Chang et al., 2010) – although this research did not investigate the specific components of trusting belief. Positive interactions with a partner in the Prisoner's Dilemma (Singer, Kiebel, Winston, Dolan, & Frith, 2004) and in the Trust Game (Delgado, Frank, & Phelps, 2005) also resulted in enhanced judgments of moral character and the corresponding changes in neurological activity. Trusting beliefs are expected to influence trust-related behaviors with the person to whom those beliefs apply. Trusting beliefs are represented in the center of the diagram in Figure 1.

Trust-related behaviors describe any actions in which a person willingly becomes vulnerable to another for a potential benefit. This includes actions traditionally associated with trust, e.g., cooperation, investment, or risk-taking. However, it can also include any actions in which a trade-off of potential benefit for risk is possible. For example, a simple interaction with another person presents the possibility of an engaging conversation at the risk of an aversive one. Engaging in conversation can be viewed as a trust-related behavior given that the risk of an aversive conversation is determined at least in part by the other party. When initially interacting with a new person, trust-related behaviors are expected to be influenced predominantly by trusting disposition. After repeated experience, trust-related behaviors are expected to change along with trusting beliefs. Previous research has found that repeated positive experience led to increased amounts of money sent to a partner in the Trust Game (Chang et al., 2010) and increased neurological markers of intentions to engage in trust-related behaviors (King-Casas et al., 2005). Trust-related behaviors are represented on the right of the diagram in Figure 1.

1.2 Trust, stereotypes, and snap judgments

The model of trust proposed by McKnight and Chervany shares much in common with research in stereotypes and group perception. In particular, research on social groups suggests that perceptions of social groups as warm and competent ultimately influence group-based prejudice or the lack thereof (Cuddy, Fiske, & Glick, 2008). Other researchers suggest that perceptions of warmth can be separated into sociability and morality; with morality being especially influential in group-directed behaviors (Brambilla, Sacchi, Rusconi, Cherubini, & Yzerbyt, 2012; Leach, Ellemers, & Barreto, 2007). This latter set of constructs maps closely onto the dimensions of competence, benevolence, and integrity that characterize faith in humanity and trusting beliefs. Indeed, while studies in interpersonal trust are often focused on the relationship between individuals rather than between groups, researchers have found that group perceptions can be transferred onto individuals that belong to those groups (McEvily, Weber, Bicchieri, & Ho, 2002). Thus, stereotypes and group perception research can be helpful in driving predictions for trust research.

The concept of stereotypes, in particular, seems to represent a group-specific form of faith in humanity. While trust research traditionally suggests that faith in humanity reflects baseline trusting beliefs of *all* others, stereotype research suggests that we may have different baseline trusting beliefs depending upon the *group membership* of others. Under this lens, researchers in group perception have found evidence that faith in humanity can influence trust-related behaviors. For example, decreased perceptions of group morality have been tied to decreased desires to interact with members of that group (Brambilla, Sacchi, Pagliaro, & Ellemers, 2013), and in-group categorization has been linked to increased cooperation (Wit & Wilke, 1992). Research in group perception also suggests that trusting beliefs can be updated

through repeated experience. Specifically, researchers have found that increased exposure to other groups increases perceptions of that group's competence, sociability, and morality (Brambilla, Hewstone, & Colucci, 2013).

Research on snap judgments based on facial appearance is also connected to trust along similar constructs. For example, researchers have found that people form competence evaluations within 100 ms of seeing another person's face (Willis & Todorov, 2006) and feelings of liking within 160 ms (Pizzagalli et al., 2002). These evaluations reflect similar dimensions of competence and benevolence that characterize faith in humanity and trusting beliefs. Moreover, the idea that judgments of competence and benevolence can be made with only minimal interaction are consistent with the concept of faith in humanity driving initial trusting beliefs. As with the stereotype and group perception research, judgments based on facial appearances have also been shown to influence trust-related behaviors. Facial appearance has been shown to be predictive of behavior in the economic Trust Game (Campellone & Kring, 2013; DeBruine, 2002; Eckel & Wilson, 2003; Scharlemann, Eckel, Kacelnik, & Wilson, 2001; van 't Wout & Sanfey, 2008) and even U.S. congressional elections (Todorov, Mandisodza, Goren, & Hall, 2005).

1.3 Current research

While previous research has found that first impressions and experience influence trust at a high-level, the present research focuses on how these two factors influence the development of trust along its various dimensions.

The first study focuses on how trusting beliefs – perceptions of the partner's competence, benevolence, and integrity – evolve from first impressions and over the course of repeated

experience. As noted, previous research has found that these components may be influenced differently by the same interventions (Kim et al., 2006; Xie & Peng, 2009) and can have different consequences on trust-related behaviors (Levin & Cross, 2004). Moreover, research in stereotypes and group perception have found that similar constructs have differing levels of importance in determining how we interact with other groups (Brambilla, Hewstone, et al., 2013; Brambilla et al., 2012; Leach et al., 2007).

The second study focuses on how trusting dispositions – faith in humanity and trusting stance – may evolve after repeated experiences. While previous research has found that trusting beliefs and trust-related behaviors can change for a single partner, it is unclear whether such changes also transfer to new partners – reflecting a possible change in trusting dispositions – or reflect only learning about that specific partner.

2. Study 1: Formation and updating of trusting beliefs

Previous research has found that trusting behaviors and high-level measures of trusting beliefs are initially set by first impressions and then influenced by repeated experience. The goal of Study 1 was to replicate and extend this research by looking at whether first impressions and experience would influence judgments of competence, benevolence, and integrity differently. The efforts in Study 1 were designed to be exploratory. While no directional predictions regarding changes in these components were made ex ante, retesting the findings of Study 1 was intended as part of the efforts for Study 2.

2.1 Methods

We adopted a similar methodology to Chang and colleagues (2010) in which participants were provided with a picture of a hypothetical partner and were then asked to engage in a repeated Trust Game with that partner.

The Trust Game (also known as the Investment Game) is a sequential game that involves two players. The first mover receives an initial allocation of money, while the second mover receives nothing. The first mover chooses how much of their initial allocation to send to the second mover, who receives triple the amount of anything that the first mover sends. The second mover then chooses how much money to send back to the first mover. The first mover has the potential to earn more money, if the second mover returns more than the initial amount sent. However, the first mover also risks losing money if the second mover sends little or no money back. The second mover also has the potential to earn more money, if the first mover sends any amount of money at all. In the traditional solution to the Trust Game, the second mover has no incentive to send any money back if they receive any, and thus should not return any money. The first mover, knowing this, should not send any money to the second mover. Nonetheless, in the original study of the Trust Game, a large majority of first movers chose to send money to their partners (Berg, Dickhaut, & McCabe, 1995).

For our studies, we used a discrete, repeated form of the Trust Game in which the first mover is provided with 2 points. The first mover could choose to send all or none of her points to the second mover. As in the standard version, if the first mover sent all of her points to the second mover, the second mover received triple the points (6 points). The second mover could

then choose to send back half of the points ("reciprocate") or none of the points ("keep"). The game was then repeated for several rounds.

2.1.1 Participants

Participants were recruited from Amazon Mechanical Turk for an "Impressions and Games" study and were compensated $1 for participation and a potential bonus based upon the outcome of the trust games. Bonuses were based on the points earned in the repeated Trust Game, which were converted at a rate of $0.05 per point. Bonuses had a final range of pay-outs from $0.45 to $3.45. A total of 391 participants completed the study. The mean age of participants was 31, and 45% of the participants were female.

2.1.2 Design

Participants played a repeated Trust Game with a simulated partner. The study used a 2x2 design, varying the partner's appearance (trustworthy or untrustworthy) and behavior (reciprocating 85% of the time or keeping 85% of the time).

Participants were first shown the picture of their partner – one of six possible photographs that they were told came from an international student website – and were asked to evaluate their partner's trustworthiness using a 12-item, 7-point Likert scale. Three of these partners had a trustworthy appearance and three of these partners had an untrustworthy appearance, as determined in an earlier pilot ($n = 555$, 37% female, $M_{age} = 29$). In the pilot, a total of 24 faces, including males and females, were rated on trustworthiness on a scale with endpoints 1 = very untrustworthy and 5 = very trustworthy. We selected three faces that were rated to be trustworthy ($M_{trust} = 3.23$, 3.38, 3.62) and three faces that were rated to be untrustworthy ($M_{trust} = 2.52$, 2.73, 2.79) for the main study. Participants then played a practice

round of the trust game, with a different "practice" partner. After the practice round, participants

played the repeated trust game with the original partner with whom they were matched. The

participant acted as the first mover with the simulated partner acting as the second mover. The

repeated Trust Game lasted 20 rounds. The partner's behavior was determined in advance of the

game, and the partner would either reciprocate for 85% of the rounds or keep the points for 85%

of the rounds. The rounds in which the partner deviated from their normal behavior were

randomly determined prior to the game. Participants were then asked to evaluate their partners

again after the game.

2.1.3 Measures

The study focused on two constructs: trust-related behavior and trusting beliefs. Trust-

related behavior was measured for each round of the repeated Trust Game and was based upon

the participant's decision of whether or not to send points to their partner. We particularly

focused on the participant's decisions during the first and last rounds to assess first impressions

and changes away from that first impression across the course of the study. Overall trusting

beliefs were measured using an 11-item, 7-point scale, including three separable subscales for

competence, benevolence, and integrity. For competence, participants were asked to indicate

how strongly they believed (strongly agree to strongly disagree) that their partner was intelligent,

skillful, and competent; for benevolence, that their partner was greedy, kind, friendly, helpful,

and nice; and for integrity, that their partner was dishonest, manipulative, and moral. Trusting

beliefs were measured before and after playing the repeated Trust Game.

2.2 Results

A summary of mean trust-related behavior and trusting beliefs are provided in Table 1. As the scale values are not inherently meaningful, trusting belief measures were standardized centered on the sample mean of each scale at the start of the game and scaled by the pooled standard deviation across the different groups. Chronbach's alphas were calculated for the overall trusting belief scale and the competence, benevolence, and integrity subscales. At the start of the game, Chronbach's alphas for the respective scales were 0.89, 0.86, 0.86, and 0.57; at the end of the game, Chronbach's alphas for the respective scales were 0.93, 0.91, 0.96, and 0.72.

To test for effects on trust-related behavior, the participant's likelihood of sending points in the first and in the last round were treated as dependent variables were regressed in two separate logistic regressions based on the partner's appearance (trustworthy or untrustworthy), behavior (reciprocate 85% of the time or keep 85% of the time), and the interaction. To test for the effects on trusting beliefs, the trusting belief scale and subscales were each regressed in two ordinary least squares regressions on the partner's appearance and behavior in one and the interaction in the other. Results for these regressions are presented in Table 2. Trust-related behavior in each round of the repeated Trust Game is presented in Figure 2.

At the start of the game, the likelihood of sending points was significantly higher for partners with a trustworthy appearance and it was not correlated with reciprocation; at the end of the game, the likelihood was significantly higher for partners who reciprocated but was not correlated with appearance. This suggested that the effect of partner appearance on trust-related behavior faded with experience.

At the start of the game, overall trusting beliefs were significantly higher for partners with a trustworthy appearance, but overall trusting beliefs were not correlated with reciprocation. At the end of the game, overall trusting beliefs showed a potential interaction between the partner's behavior and appearance. Overall trusting beliefs in partners who reciprocated were similar for both trustworthy and untrustworthy appearances, and these beliefs were higher than those in partners who kept the points. However, overall trusting beliefs in partners who kept the points and had an untrustworthy appearance were even lower than beliefs in partners who kept the points and had a trustworthy appearance.

For competence, we found a similar pattern of significance in the first and last rounds as trusting beliefs. At the start of the game, competence beliefs were significantly higher for partners with a trustworthy appearance and were not correlated with reciprocation. At the end of the game, competence beliefs were similarly high for partners who reciprocated regardless of appearance, lower for partners who kept the points, and particularly low for those who kept the points and had an untrustworthy appearance. For benevolence and integrity, we found a similar pattern of significance in the first and last rounds as trust-related behavior. At the start of the game, benevolence and integrity beliefs were significantly higher for partners with a trustworthy appearance and were not correlated with reciprocation; but at the end, benevolence and integrity beliefs were significantly higher for reciprocating partners and not just for those with a trustworthy appearance.

2.3 Discussion

2.3.1 Trusting beliefs

A high-level view of the results from Study 1 suggest that overall trusting beliefs are higher at the start of the game for partners with a trustworthy appearance and higher at the end of the game for partners who reciprocate. While overall trusting beliefs at the end of the game included an interaction effect – a comparison of the means in Table 1 suggests that the effect of partner reciprocation was much stronger than the residual effects of the partner's appearance on those partners who kept points. These results are similar to earlier studies, which found a strong influence of partner appearance at the start of the game, a strong influence of partner behavior at the end of the game, and a residual influence of partner appearance for those who didn't reciprocate (Chang et al., 2010).

Looking at the components of overall trusting belief, we found that competence followed a similar pattern as overall trusting beliefs (with an interaction effect at the end of the game) whereas benevolence and integrity did not. This suggests that the interaction effect observed in overall trusting beliefs may be attributable to competence judgments, rather than benevolence or integrity judgments. On the one hand, the interaction effect in competence may represent a lasting influence of partner appearance on judgments of competence even after experience. However, an alternate explanation is that competence beliefs update more slowly than benevolence and integrity beliefs, and the interaction effect is an artifact of stopping the repeated Trust Game before competence beliefs were fully adapted. Note that under the reinforcement learning model proposed by Chang et al. (2010), trusting beliefs should update gradually from an initial value based on partner appearance to a longer-term value based on partner behavior.

Following from these assumptions, we would expect the following pattern of results. (1) Partner appearance determines trusting beliefs at the very start of the repeated Trust Game. (2) After the start, but early in the repeated Trust Game, partner appearance and partner behavior independently influence trusting beliefs, resulting in two main effects. (3) Later in the repeated Trust Game, we should see an interaction effect as trusting beliefs converge on one of the long-term partner behavior trust levels before the other. (4) Finally, after many rounds, we should see a main effect of trusting beliefs only. Looking at trust-related behaviors as a proxy for trusting beliefs, Figure 2 shows a pattern of results similar to what would be predicted – with a main effect of appearance in the first round (1), indications of a possible interaction effect in the third round (3), and a main effect of behavior in the final round (4). The lack of two independent main effects (2) may represent a quick adaptation of trust-related behaviors. If trusting beliefs follow a similar pattern of gradual updating, the interaction effect in competence beliefs and the main effect of behavior only for benevolence and integrity beliefs may indicate that competence beliefs were earlier in its updating process (3) than either benevolence or integrity (4) when the repeated Trust Game was stopped. As such, this suggests that experience in the repeated Trust Game influences perceived competence more slowly than it influences perceived benevolence and integrity. These findings are consistent with stereotype research suggesting that compared to competence, benevolence and integrity are particularly important during initial impression formation (Brambilla et al., 2012; Cuddy et al., 2008).

2.3.2 Trust-related behavior

In the first round of the repeated Trust Game, participants' decisions to trust the partner depended strongly on the partner's appearance; in the last round, decisions depended strongly on the partner's behavior. These findings are consistent with the broader trust model. First, the link

between appearance, trusting beliefs, and sharing matches expectations that trusting dispositions influence initial trusting beliefs, which then influence trust-related behavior. Second, the change in trusting beliefs and trust-related behaviors between the start and end matches expectations that trusting beliefs for a specific partner are revised after repeated interaction and that these changes bring about changes in behavior after time. These findings are consistent with previous research (Chang et al., 2010), finding similar effects of partner appearance on the first round and effects of partner behavior on the final round trust-related behavior. However, the earlier findings also find an interaction effect in the final rounds, whereas our study finds none. As this earlier study involved fewer rounds than the present study (15 vs. 25), these results may be best explained through differences in the degree to which trusting beliefs have stabilized.

Nonetheless, inconsistencies regarding interaction effects between the present and previous studies raise concerns about their robustness. Furthermore, we are interested in learning whether the participant's experiences in the repeated Trust Game simply updated trusting beliefs about the specific partner or if those experiences may have adjusted the participant's faith in humanity. These ideas were tested in Study 2.

3. Study 2: Developing trusting dispositions

In Study 1, we found that, with experience, both the participant's trusting beliefs and trust-related behaviors changed in a manner consistent with the partner's behavior and became less dependent upon the partner's appearance. However, as these participants were paired with only one partner, it is difficult to determine whether such results reflected changes only in how the participant viewed this specific partner; or whether such results reflected changes in how the participant viewed others more broadly, i.e., trusting dispositions.

In Study 2, in addition to studying how participants behaved at the end of a single repeated Trust Game, we also turn our attention to how participants behaved when paired with a new partner *after* a single repeated Trust Game. If participants only changed their trusting beliefs in the partner they were paired with, we should expect that participants playing with a new partner will continue to use the appearance of their new partner to drive their initial trusting beliefs and initial trust-related behavior. However, if participants' experiences in the first repeated Trust Game revised their faith in humanity, we may find that participants are less likely to use the new partner's appearance to determine initial trusting beliefs and initial trust-related behavior.

To the degree that faith in humanity might be appearance-specific – as is suggested by the literature on stereotypes and on evaluations of faces – we might also expect that *changes* in faith in humanity might also be appearance-specific. Thus, we propose two alternate hypotheses for how faith in humanity might be revised after a single experience in the repeated Trust Game. If participants change the way they approach all others, we refer to this as changes in general faith in humanity. If the participants change the way they approach only the "type" of person that they observed in the first experience of the repeated Trust Game, we refer to this as changes in appearance-specific faith in humanity. The underlying null hypothesis is that faith in humanity is not changed at all. The two alternate hypotheses are discussed in more detail, below.

Hypothesis 1. Changes in general faith in humanity. Repeated interaction in the Trust Game in which the partner consistently reciprocates may increase a participant's general faith in humanity – leading her to approach *all* new partners with greater beliefs in their competence, benevolence, and integrity. Likewise, repeated interactions with a partner who consistently keeps

points sent in a Trust Game may lead to decreases in a participant's general faith in humanity,

resulting in decreased beliefs in new partners' competence, benevolence, and integrity.

In this case, participants previously matched with a partner that reciprocated would

demonstrate increased trust-related behavior and report higher levels of trusting belief in a

second partner. Likewise, participants previously matched with a partner that kept any points

sent should demonstrate decreased trust-related behavior and report lower levels of trusting

belief.

Hypothesis 2. Changes in appearance-specific faith in humanity. As discussed, stereotype

research suggests that faith in humanity may vary depending upon the group to which a partner

belongs, which may in turn be determined by that partner's appearance. If a participant is

previously matched with a partner who appears trustworthy and reciprocates or who appears

untrustworthy and keeps the points, this experience may reaffirm the participant's prior

stereotypes. Participants may then depend more heavily upon the appearance of a new partner in

constructing their trusting beliefs. If a participant is previously matched with a partner who

appears untrustworthy and reciprocates or who appears trustworthy and keeps the points, this

may disprove the stereotypes that the participant held, leading to less reliance upon the partner's

appearance in forming initial trusting beliefs.

In the context of the Trust Game, this would result in four predictions. Compared to

participants with no previous experience, participants who have had their stereotypes reaffirmed

may report higher initial trusting beliefs and engage in more trust-related behaviors when

matched with a new partner with a trustworthy appearance and lower initial trusting beliefs and

fewer trust-related behaviors when matched with a new partner with an untrustworthy

appearance. Likewise, participants who had their stereotypes disproven may report lower initial

trusting beliefs and engage in fewer trust-related behaviors when matched with a new partner

with a trustworthy appearance and higher initial trusting beliefs and more trust-related behaviors

when matched with a new partner with an untrustworthy appearance.

3.1 Methods

3.1.1 Participants

Participants were recruited from Amazon Mechanical Turk for an "Impressions and

Games" study and were compensated $2 for participation and offered a potential bonus for

performance. Participants were paid a bonus based on their performance across two Trust

Games, which were converted at a rate of $0.03 per point and with a final range of pay-outs from

$0 to $2.70. A total of 265 participants completed the study. The mean age of participants was

31, and 50% of the participants were female.

3.1.2 Design

In Study 2, participants played the trust game with two different partners. The study used

a 2x2x2 between-subjects design, varying the first partner's appearance (trustworthy or

untrustworthy), the first partner's behavior (reciprocate or keep), and the second partner's

appearance (trustworthy or untrustworthy).

Participants were first shown the picture of one of four possible partners – two of these

partners had a trustworthy appearance and two had an untrustworthy appearance, as determined

by an earlier pilot (n = 50, 60% female, mean age = 26). Participants were then asked to evaluate

their partner based upon these pictures. To test for the robustness of Study 1's effects, these

pictures differed from those used in Study 1. To reduce noise caused by anything outside of

facial stereotypes, all pictures selected were of white, middle-aged males with clothing and

background digitally covered. Participants then played a practice round of the Trust Game with a

different "practice" partner. After the practice round, participants played the repeated Trust

Game with the original partner with whom they were matched. The participant acted as the first

mover, with the simulated partner acting as the second mover. The repeated Trust Game lasted

15 rounds (fewer than Study 1). The partner's behavior was determined in advance of the game,

and the partner would either always reciprocate or always keep the points. The participants were

then asked to evaluate their partners a second time after the game. Participants were then shown

a picture of a different person, one of the three remaining possible partners, and repeated the

process with the new partner.

The picture shown to the participant was counterbalanced such that participants were

equally likely to receive a partner with a trustworthy or untrustworthy appearance for both the

first and second repeated trust games. Due to technical errors, the second partner's behavior

always matched that of the first partner. However, as the focus of Study 2 was on trust-related

behavior and beliefs in the first game and trust-related behavior and beliefs at the start of the

second game, before the second partner responded, this error did not affect our analysis.

Participants were informed that the partners were hypothetical, but that their responses in the

repeated Trust Game were based on previously observed behavior.

3.1.3 Measures

Study 2 focused on three measures: faith in humanity, trust-related behavior (including

the subcomponents of competence, benevolence, and integrity), and trusting beliefs. Faith in

humanity was measured by looking for changes in trusting beliefs in the second game resulting

from experiences in the first game. Overall trusting beliefs were measured at the start and at end

of each repeated Trust Game using a 9-item, 7-point scale, including three separable subscales

for competence, benevolence, and integrity. For competence, participants were asked to indicate

how strongly they believed (strongly agree to strongly disagree) that their partner was intelligent,

skillful, incompetent; for benevolence, that their partner was greedy, kind, friendly; and for

integrity, that their partner was honest, reliable, manipulative. During the first three and last three

rounds, participants were additionally asked whether or not they believed their partner would

reciprocate if they were sent points. Trust-related behavior was measured for each round of the

repeated Trust Game and was based upon the participant's decision of whether or not to send

points to their partners.

3.2 Results

3.2.1 Replication of Study 1

A summary of mean trust-related behavior and trusting beliefs at the start and end of

Game 1, including subscales for competence, benevolence, and integrity, are provided in Table

3. The overall trusting belief scale and its components were standardized based upon the mean of

trusting beliefs at the start of the game and scaled by the standard deviation pooled by whether

the first partner had a trustworthy or an untrustworthy appearance. Chronbach's alphas were

calculated for the aggregate trustworthiness scale and the competence, benevolence, and integrity

subscales. At the start of the first game, Chronbach's alphas for the respective scales were 0.92,

0.80, 0.83, and 0.78; at the end of the first game, they were 0.93, 0.80, 0.91, and 0.87; at the start

of the second game, they were 0.93, 0.78, 0.84, 0.82; and at the end of the second game, they were 0.94, 0.83, 0.90, 0.89.

Regressions of trust-related behavior and trusting beliefs are presented in Table 4. At the start of Game 1, the likelihood of sending points is significantly higher for partners with a trustworthy appearance and is not correlated with reciprocation; at the end of the game, the likelihood is significantly higher for partners who reciprocate but is not correlated with appearance.

At the start of Game 1, overall trusting beliefs are significantly higher for partners with a trustworthy appearance and are not correlated with reciprocation. At the end of the game, overall trusting beliefs are higher when the partner has a trustworthy appearance and when the partner reciprocated, with no significant interaction.

Competence beliefs are significantly higher at the start of the game for partners with a trustworthy appearance and are not correlated with reciprocation. At the end of the game, competence beliefs are higher when the partner has a trustworthy appearance and when the partner reciprocated, with no significant interaction. Benevolence and integrity beliefs are also significantly higher at the start of the game for partners with a trustworthy appearance and are not correlated with reciprocation. At the end of the game, benevolence and integrity beliefs involve a significant interaction: both beliefs were lower for partners who kept the points and higher for partners who reciprocated, but benevolence and integrity beliefs in partners who reciprocated and had a trustworthy appearance were significantly higher than benevolence and integrity beliefs in partners who reciprocated and had an untrustworthy appearance.

3.2.2 Trusting disposition

Table 5 provides mean values for trust-related behavior (% participants sharing) and trusting beliefs (scale and beliefs about partner's behavior) at the start of the first and second repeated Trust Games. For both Game 1 and Game 2, these are grouped by their current partner's appearance. For Game 2, these are also grouped by their first partner's appearance and behavior. The trusting beliefs scale is standardized using the mean and pooled standard deviation (by treatment) of the evaluations from the start of Game 1.

To test for changes in behavior resulting from changes in general faith in humanity, a logistic regression was run using the participant's likelihood of sending points in the first round of Game 2 as the dependent variable and the partner's appearance in Game 1, the partner's behavior in Game 1, and the partner's appearance in Game 2 as independent variables. To test for changes in behavior resulting from changes in appearance-specific faith in humanity, additional logistic regressions were run, adding a two-way interaction between the appearance and the behavior of the partner in Game 1; and then adding a three-way interaction between the appearance and the behavior of the partner in Game 1 and the appearance of the partner in Game 2. These regressions are included in Table 6 as models 1A, 1B, and 1C, respectively.

To test for changes in general faith in humanity directly, an ordinary least squares regression was run using the trusting beliefs scale at the start of Game 2 as the dependent variable and the partner's appearance in Game 1, the partner's behavior in Game 1, and the partner's appearance in Game 2 as independent variables. To test for changes in appearance-specific faith in humanity, additional regressions were run adding a two-way and three-way interaction term. These regressions are included in Table 6 as models 2A, 2B, and 2C,

respectively. Regressions for the trusting belief components maintained the same significance

ratings of the overall trusting belief scale, so they are not included in the table.

Additionally, a logistic regression was run using the participant's belief that their partner

would reciprocate in the first round of Game 2 as the dependent variable and the partner's

appearance in Game 1, the partner's behavior in Game 1, and the partner's appearance in Game 2

as independent variables. Again, regressions including a two-way and three-way interaction term

were included. These regressions are included in Table 6 as models 3A, 3B, and 3C,

respectively.

Models 1A, 2A, and 3A indicate that the partner's behavior in Game 1 (P1 Reciprocated)

does not have a significant effect on trust-related behavior, trusting beliefs, or beliefs in

reciprocation. Overall, there was no evidence for changes in general faith in humanity.

Models 1B, 2B, and 3B show that the interaction term of Game 1's partner's appearance

and behavior (P1 Trustworthy x P1 Reciprocated) has no significant effect on trust-related

behavior, trusting beliefs, or beliefs in reciprocation. This lack of significance is also reflected in

likelihood ratio tests comparing the models to Models 1A, 2A, and 3A – for trust-related

behavior, $\chi^2(1) = 0.08$, $p = 0.78$, trusting beliefs, $\chi^2(1) = 0.90$, $p = 0.34$, and beliefs in

reciprocation, $\chi^2(1) = 1.35$, $p = 0.25$. This suggests no evidence of changes in appearance-

specific faith in humanity.

Reaffirmed or disconfirmed stereotypes about trustworthy faces might not affect

stereotypes of untrustworthy faces. Models 1C, 2C, and 3C include a three-way interaction

allowing for this possibility. These models indicate that the three-way interaction has no

significant effect on trust-related behavior, trusting beliefs, or beliefs in reciprocity Likelihood

ratio tests comparing Models 1C, 2C, and 3C against Models 1A, 2A, and 3A indicate that the two-way and three-way interactions are jointly non-significant for trust-related behavior, $\chi^2(2) = 0.10$, $p = 0.94$, trusting beliefs, $\chi^2(1) = 0.94$, $p = 0.62$, and beliefs in reciprocity, $\chi^2(2) = 1.35$, $p = 0.50$. Again, there is no evidence of changes in appearance-specific faith in humanity.

3.2.3 Additional post-hoc comparisons

Post-hoc observations in Table 5 suggest additional differences of note. The likelihood of sending points to a partner with an untrustworthy appearance is higher in Game 2 compared to Game 1, $M = 30.29$, $t(263)$, $p < 0.001$. This difference is not fully reflected in beliefs in reciprocity. Notably, the proportion of participants sending points believing that the partner will *not* return them increases in Game 2 compared to Game 1, $M = 0.10$, $t(528)$, $p = 0.002$.

3.3 Discussion

3.3.1 Replication of Study 1

Study 2 successfully replicates the effects from Study 1, with a significant effect of reciprocation increasing trust-related behavior in the last round and no significant effect of the partner's appearance. Study 2 also replicates a significant effect of reciprocation increasing trusting beliefs at the end of the repeated Trust Game; however, it also presents evidence of a significant interaction effect for benevolence and integrity beliefs and a significant main effect of partner's appearance for competence. Nonetheless, the consistently strong effect size for trusting beliefs suggests a robust effect of reciprocation in later rounds of the repeated Trust Game.

While differences in interaction effects between Study 1 and Study 2 may have arisen from contextual differences, the pattern of results also remains consistent with the explanation

that competence beliefs update more slowly with experience than benevolence and integrity beliefs. Note that Study 2 ended after fewer rounds than Study 1, such that the updating of competence, benevolence, and integrity beliefs are likely to have been interrupted sooner than they were in Study 1. Finding a main effect of both partner appearance and behavior for competence is suggestive of competence being in the early phases of updating, whereas finding an interaction effect of partner appearance and behavior is suggestive of benevolence and integrity being in a later phase of updating. As such, in spite of the different significance findings, Study 2 nonetheless provides evidence that is consistent with Study 1 – suggesting that competence updates more slowly than benevolence and integrity.

3.3.2 Trusting disposition

Study 2 suggests that experiences in a repeated Trust Game played with a single partner do not appear to significantly affect a participant's faith in humanity – we continue to use appearance as a basis for forming initial trusting beliefs in a new partner. As such, the continued impact of first impressions may be more underappreciated when we only look at a single repeated Trust Game. As a corollary, the failure to find a significant change in faith in humanity within Study 2 should also not be over-interpreted. It is possible that after *several* repeated Trust Games, participants may eventually revise their faith in humanity. Nonetheless, this still suggests that a substantial effort may be required to change what our baseline beliefs are.

Post-hoc analysis of Study 2, however, suggests that trusting disposition may still be sensitive to a single instance of a repeated Trust Game. In particular, we find that when starting Game 2, participants were more likely to send points to their partner, even when they did not expect their partner to reciprocate – that is, they engaged in trust-related behaviors despite

having low trusting beliefs. One possible explanation for this behavior is that participants are adopting a trusting stance and are choosing to engage in trust-related behaviors for strategic reasons. In particular, the design of the repeated Trust Game is such that the first mover only receives information about whether the participant will reciprocate if the first mover engages in some level of trust-related behavior. Sending points in the first round when the expected return is negative can then be interpreted as a "purchase" of information that allows for more informed decisions later on.

4. Conclusions

As in previous studies, we found greater trusting beliefs and more trust-related behavior for partners with a trustworthy facial appearance at the start of the repeated Trust Game, and for partners who reciprocated at the end of the repeated Trust Game. However, the present research provides a more comprehensive understanding of the influence of these factors on different dimensions of trust. First, we confirmed that first impressions and experience have similar effects across the different dimensions of trusting belief – competence, benevolence, and integrity. However, we noted suggestive evidence that relative to benevolence and integrity beliefs, experience in the repeated Trust Game may be slower to affect competence beliefs. Second, we found that a single experience in the repeated Trust Game does not affect our reliance on facial appearance in forming initial trusting beliefs, even when our previous expectations based on facial appearance were disconfirmed. This implies that a single disconfirming experience is unlikely to influence our general beliefs in the trustworthiness of others (faith in humanity), even while it may influence our specific beliefs in the trustworthiness of a particular individual. Finally, we found that a single experience in the repeated Trust Game can influence trusting-behavior independent of trusting beliefs. In particular, participants were

more likely to share with their partner, even when they did not expect their trust to be reciprocated. This implies that people may be quicker to adapt their strategic approach towards trust *situations* in general (trusting stance), even while they do not adapt their beliefs in the trustworthiness of other *people* in general.

From a theoretical perspective, our studies improve our understanding of how different dimensions of trust evolve as a result of first impressions and experience. In some cases, these dimensions move consistently, e.g., trusting behaviors appear to reflect trusting beliefs for a single partner over time. In others, they move independently, e.g., trusting behaviors *do not* reflect trusting beliefs in early interactions with a second partner. This specification also allows for a clearer understanding of *what* may be learned in repeated interactions in the Trust Game. In particular, we find that changes in trust-related behavior reflect changes in trusting beliefs relating to a specific individual rather than changes in trusting beliefs about people in general.

This methodological approach can also help in the interpretation of previous research. For example, Campellone and Kring (2013) study how facial expressions of emotion influence trust judgments and behavior in the Trust Game. They find that participants judge angry faces and non-angry faces to be similar in "trustworthiness"; however, they are more likely to share with non-angry faces than with angry faces. By breaking down "trustworthiness" into specific components of warmth, competence, and integrity, these results might be easier to reconcile. Angry faces, while low in benevolence, are also fairly high in integrity. While these faces suggest that they are unlikely to reciprocate, they express their intent fairly explicitly. These two effects could counteract each other, resulting in non-significant differences in judgments of overall trustworthiness. On the other hand, because these angry faces so clearly communicate

their intention not to reciprocate, it's unsurprising that participants would be less likely to share with them.

From a more applied perspective, our research suggests several complications in the area of trust management. Our findings that experience influences trusting beliefs regarding specific individuals rather than people in general suggest that it may be difficult to promote trusting beliefs more broadly. For a manager in an organization low on trust, leading by example may be insufficient, as it may simply build trust in the manager but not across employees. Connecting general trust in others to group stereotypes, this also suggests that providing a positive example of an out-group member is unlikely to reduce the use of stereotypes for other members of the same out-group. Indeed, researchers have found that out-group members that disconfirm stereotypes are often subtyped so that the initial out-group stereotype can be maintained (Richards & Hewstone, 2001). As a consequence, improving trust across individuals more broadly can be more challenging than it may appear.

On the other hand, our findings that participants are willing to engage in trust-related behaviors independent of trusting beliefs suggest a potential process for building trust more broadly. Interventions designed to promote the strategic advantages of trust – rather than to enhance perceptions of the trustworthiness of others – have the potential to promote early trust-related behaviors. If these early behaviors are reciprocated, enhanced trusting beliefs can develop through experience. Indeed, this type of intervention might be found in the Russian adage "trust, but verify" – a phrase often used by Ronald Reagan to Mikhail Gorbachev as the two continued to work towards easing tensions between their respective countries.

In this paper, we clarify the effects of first impressions and experience on trust in the context of the repeated Trust Game. Looking forward, it will be important to consider how well our results may hold in other contexts. For example, while we find evidence that competence beliefs may update more slowly than benevolence and integrity beliefs, these findings may be particular to the repeated Trust Game, in which competence is an unlikely explanation for a failure to reciprocate. Alternate contexts, might find that competence judgments may adapt more quickly – for example, if the partner is asked to solve a problem on behalf of the participant. In our studies, we also find that after a single experience in the repeated Trust Game, participants continue to rely on appearances to form initial trusting beliefs; however, it's not clear whether this reliance is robust after *several* previous experiences in the repeated Trust Game. Finally, while participants' use of appearances to drive initial trust-related behaviors (faith in humanity) seems hard to change, their decisions to engage in trust-related behaviors for strategic reasons (trusting stance) may be more adaptable. Research that tests the relative effectiveness of interventions focused on improving faith in humanity and interventions focused on improving a trusting stance may help us to better understand how we might increase trust more broadly.

References

Ben-Ner, A., & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, *31*(1), 64–79. doi:10.1016/j.joep.2009.10.001

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142. doi:10.1006/game.1995.1027

Brambilla, M., Hewstone, M., & Colucci, F. P. (2013). Enhancing moral virtues: Increased perceived outgroup morality as a mediator of intergroup contact effects. *Group Processes & Intergroup Relations*, *16*(5), 648–657. doi:10.1177/1368430212471737

Brambilla, M., Sacchi, S., Pagliaro, S., & Ellemers, N. (2013). Morality and intergroup relations: Threats to safety and group image predict the desire to interact with outgroup and ingroup members. *Journal of Experimental Social Psychology*, *49*(5), 811–821. doi:10.1016/j.jesp.2013.04.005

Brambilla, M., Sacchi, S., Rusconi, P., Cherubini, P., & Yzerbyt, V. Y. (2012). You want to give a good impression? Be honest! Moral traits dominate group impression formation. *British Journal of Social Psychology*, *51*(1), 149–166. doi:10.1111/j.2044-8309.2010.02011.x

Campellone, T. R., & Kring, A. M. (2013). Who do you trust? The impact of facial emotion and behaviour on decision making. *Cognition & Emotion*, *27*(4), 603–20. doi:10.1080/02699931.2012.726608

Chang, L. J., Doll, B. B., van 't Wout, M., Frank, M. J., & Sanfey, A. G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology*, *61*(2), 87–105. doi:10.1016/j.cogpsych.2010.03.001

Cuddy, A., Fiske, S., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 40, pp. 61–149). New York, NY: Academic Press. doi:10.1016/S0065-2601(07)00002-0

DeBruine, L. M. (2002). Facial resemblance enhances trust. *Proceedings of the Royal Society of London, Series B*, *269*(1498), 1307–12. doi:10.1098/rspb.2002.2034

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*(11), 1611–1618. doi:10.1038/nn1575

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but ...: A meta-analytic review of research on the physical attractiveness of stereotype. *Psychological Bulletin*, *110*(1), 109–128. doi:10.1037/0033-2909.110.1.109

Eckel, C. C., & Wilson, R. K. (2003). The human face of game theory: Trust and reciprocity in sequential games. In E. Ostrom & J. Walker (Eds.), *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research* (pp. 245–274). New York, NY: Russel Sage Foundation.

Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin*, *111*(2), 304–341. doi:10.1037/0033-2909.111.2.304

Hamermesh, D. S., & Biddle, J. E. (1994). Beauty and the labor market. *American Economic Review*, *84*(5), 1174–1194.

Kim, P. H., Dirks, K. T., Cooper, C. D., & Ferrin, D. L. (2006). When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence- vs. integrity-based trust violation. *Organizational Behavior and Human Decision Processes*, *99*(1), 49–65. doi:10.1016/j.obhdp.2005.07.002

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*(5718), 78–83. doi:10.1126/science.1108062

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology*, *93*(2), 234–249. doi:10.1037/0022-3514.93.2.234

Levin, D. Z., & Cross, R. (2004). The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management Science*, *50*(11), 1477–1490. doi:10.1287/mnsc.1030.0136

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734. doi:10.5465/AMR.1995.9508080335

McEvily, B., Weber, R. A., Bicchieri, C., & Ho, V. (2002). Can groups be trusted? An experimental study of collective trust. *Department of Social and Decision Sciences*, Paper 25.

McKnight, D. H., & Chervany, N. L. (2001). Trust and distrust definitions: One bite at a time. In R. Falcone, M. Singh, & Y.-H. Tan (Eds.), *Trust in Cyber-societies: Integrating the Human and Artifical Perspectives* (pp. 27–54). Berlin: Springer.

McKnight, D. H., & Chervany, N. L. (2002). What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology. *International Journal of Electronic Commerice*, *6*(2), 35–59.

Pizzagalli, D. A., Lehmann, D., Hendrick, A. M., Regard, M., Pascual-Marqui, R. D., & Davidson, R. J. (2002). Affective judgments of faces modulate early activity (∼160 ms) within the fusiform gyri. *NeuroImage*, *16*(3), 663–677. doi:10.1006/nimg.2002.1126

Richards, Z., & Hewstone, M. (2001). Subtyping and subgrouping: Processes for the prevention and promotion of stereotype change. *Personality and Social Psychology Review*, *5*(1), 52–73. doi:10.1207/S15327957PSPR0501

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*(5), 617–640. doi:10.1016/S0167-4870(01)00059-9

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. doi:10.1016/S0896-6273(04)00014-5

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*(5728), 1623–1626. doi:10.1126/science.1110589

Van 't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, *108*(3), 796–803. doi:10.1016/j.cognition.2008.07.002

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. doi:10.1111/j.1467-9280.2006.01750.x

Wit, A. P., & Wilke, H. A. M. (1992). The effect of social categorization on cooperation in three types of social dilemmas. *Journal of Economic Psychology*, *13*(1), 135–151. doi:10.1016/0167-4870(92)90056-D

Xie, Y., & Peng, S. (2009). How to repair customer trust after negative publicity: The roles of competence, integrity, benevolence, and forgiveness. *Psychology & Marketing*, *26*(7), 572–589. doi:10.1002/mar.20289

Zebrowitz, L. (1999). *Reading faces: Window to the soul?* Boulder, CO: Westview Press.

Table 1. Study 1: Mean (standard error) of trust-related behavior and trusting beliefs at start and end

Appearance:	Trustworthy				Untrustworthy			
Trust-related behavior:	Keep		Reciprocate		Keep		Reciprocate	
Timing:	Start	End	Start	End	Start	End	Start	End
n	98		106		93		94	
Sharing, % Participants	81.63 (3.93)	16.33 (3.75)	75.47 (4.20)	62.26 (4.73)	67.74 (4.87)	16.13 (3.83)	61.70 (5.04)	67.02 (4.88)
Trusting Beliefs	0.30 (0.09)	-1.38 (0.12)	0.29 (0.09)	0.36 (0.11)	-0.40 (0.11)	-1.86 (0.13)	-0.25 (0.10)	0.39 (0.12)
Competence	*0.27 (0.09)*	*-0.16 (0.14)*	*0.23 (0.09)*	*0.36 (0.11)*	*-0.34 (0.11)*	*-0.83 (0.16)*	*-0.20 (0.11)*	*0.33 (0.12)*
Benevolence	*0.29 (0.09)*	*-1.88 (0.13)*	*0.33 (0.09)*	*0.38 (0.11)*	*-0.43 (0.10)*	*-2.17 (0.13)*	*-0.25 (0.10)*	*0.40 (0.12)*
Integrity	*0.23 (0.10)*	*-1.32 (0.13)*	*0.16 (0.09)*	*0.19 (0.12)*	*-0.24 (0.11)*	*-1.63 (0.13)*	*-0.18 (0.10)*	*0.30 (0.13)*

Note. Trusting beliefs and components are standardized using the means and standard deviations at the start of the game. Standard deviations are pooled by partner appearance and partner behavior.

Table 2. Study 1: Regressions of trust-related behavior and trusting beliefs at start and end

DV:	Trust-related behavior		Trusting Beliefs		Competence		Benevolence		Integrity	
	Start	End	Start	End	Start	End	Start	End	Start	End
Model Type:	Logistic				Ordinary Least Squares					
Appearance (Trustworthy)	0.75 * (0.34)	0.01 (0.39)	0.70 *** (0.14)	0.47 ** (0.17)	0.61 *** (0.14)	0.66 *** (0.19)	0.71 *** (0.14)	0.30 (0.18)	0.48 ** (0.14)	0.31 (0.18)
Reciprocated	-0.27 (0.31)	2.36 *** (0.36)	0.15 (0.14)	2.25 *** (0.18)	0.14 (0.14)	1.16 *** (0.19)	0.18 (0.14)	2.57 *** (0.18)	0.06 (0.14)	1.94 *** (0.18)
Appearance x Reciprocated	-0.10 (0.46)	-0.22 (0.49)	-0.16 (0.19)	-0.51 * (0.24)	-0.18 (0.19)	-0.63 * (0.27)	-0.14 (0.19)	-0.31 (0.25)	-0.13 (0.20)	-0.42 (0.25)
Constant	0.74 *** (0.22)	-1.65 *** (0.28)	-0.40 *** (0.10)	-1.86 *** (1.24)	-0.34 *** (0.10)	-0.83 *** (0.14)	-0.43 *** (0.10)	-2.17 *** (0.13)	-0.24 (0.10)	-1.63 *** (0.13)
Adjusted R^2			0.092	0.41	0.06	0.11	0.10	0.49	0.04	0.32

Note. Numbers in parentheses are standard errors. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Trusting beliefs and components are standardized using the means and standard deviations at the start of the game. Standard deviations are pooled by partner appearance and partner behavior.

Table 3. Study 2, Game 1: Mean (standard error) of trust-related behavior, trusting beliefs, and reciprocating beliefs at start and end

Appearance:	Trustworthy				Untrustworthy			
Behavior:	Keep		Reciprocate		Keep		Reciprocate	
Timing:	Start	End	Start	End	Start	End	Start	End
n	76		62		66		61	
Sharing, % participants	88.16 (3.73)	25.00 (5.00)	90.32 (3.79)	88.71 (4.05)	51.52 (6.20)	12.12 (4.05)	42.62 (6.38)	85.25 (4.58)
Trusting beliefs	1.01 (0.11)	-1.17 (0.12)	1.03 (0.12)	1.89 (0.14)	-1.03 (0.13)	-1.86 (0.14)	-1.18 (0.13)	0.69 (0.17)
Competence	*0.74 (0.11)*	*0.05 (0.13)*	*0.70 (0.12)*	*1.25 (0.13)*	*-0.76 (0.14)*	*-1.19 (0.17)*	*-0.81 (0.13)*	*0.25 (0.15)*
Benevolence	*1.03 (0.10)*	*-1.83 (0.14)*	*1.05 (0.12)*	*1.85 (0.13)*	*-1.05 (0.14)*	*-2.11 (0.12)*	*-1.22 (0.14)*	*0.67 (0.18)*
Integrity	*0.84 (0.11)*	*-1.29 (0.14)*	*0.92 (0.12)*	*1.80 (0.14)*	*-0.87 (0.13)*	*-1.53 (0.14)*	*-1.05 (0.13)*	*0.88 (0.17)*
Belief reciprocating, % participants	86.84 (3.90)	21.05 (4.71)	93.55 (3.15)	93.55 (3.15)	43.63 (6.16)	9.09 (3.57)	36.07 (6.20)	85.25 (4.58)

Note. Trusting beliefs and components are standardized using the means and standard deviations at the start of the game. Standard deviations are pooled by partner appearance and partner behavior.

Table 4. Study 2, Game 1: Regression of trust-related behavior and trusting beliefs at start and end

Model Type:	Logistic		Ordinary Least Squares							
DV:	Sharing Behavior		Trusting beliefs		Competence		Benevolence		Integrity	
	Start	End	Start	End	Start	End	Start	End	Start	End
Appearance (Trustworthy)	1.95 *** (0.43)	0.88 (0.46)	2.04 *** (0.17)	0.69 *** (0.20)	1.50 *** (0.17)	1.23 *** (0.20)	2.09 *** (0.17)	0.28 (0.20)	1.71 ** (0.17)	0.25 (0.20)
Reciprocated	-0.36 (0.36)	3.74 *** (0.52)	-0.15 (0.17)	2.56 *** (0.21)	-0.05 (0.18)	1.44 *** (0.21)	-0.16 (0.18)	2.79 *** (0.21)	-0.87 (0.12)	2.42 *** (0.22)
Appearance x Reciprocated	-0.58 (0.66)	-0.58 (0.71)	0.17 (0.25)	0.50 (0.29)	-0.05 (0.18)	-0.24 (0.29)	0.18 (0.25)	0.90 ** (0.29)	0.26 (0.25)	0.67 * (0.30)
Constant	0.06 *** (0.25)	-1.98 *** (0.38)	-1.03 *** (0.12)	-1.86 *** (0.14)	-0.76 *** (0.25)	-1.19 *** (0.11)	-1.05 *** (0.12)	-2.11 *** (0.14)	-0.87 (0.12)	-1.54 *** (0.15)
Adjusted R^2			0.53	0.61	0.36	0.34	0.54	0.67	0.46	0.57

Note. Numbers in parentheses are standard errors. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Trusting beliefs and components are standardized using the means and standard deviations at the start of the game. Standard deviations are pooled by partner appearance and partner behavior.

Table 5. Study 2, Game 2: Mean (standard error) of trust-related behavior, trusting beliefs, and reciprocating beliefs at start

Game:	First	Second				First	Second			
First Partner Appearance:		Trustworthy					Untrustworthy			
First Partner Response:	N/A	Keep	Return	Keep	Return	N/A	Keep	Return	Keep	Return
Second Partner Appearance:	N/A	Trustworthy		Untrustworthy		N/A	Trustworthy		Untrustworthy	
n	*138*	*31*	*27*	*45*	*35*	*127*	*36*	*33*	*30*	*28*
Sharing, % Participants	89.13 (2.66)	90.32 (5.40)	88.89 (6.16)	80.00 (6.03)	74.29 (7.50)	47.24 (4.45)	94.44 (3.87)	90.91 (5.08)	76.67 (7.85)	78.57 (7.90)
Trusting beliefs	0.69 (0.05)	0.60 (0.14)	0.31 (0.17)	-0.51 (0.15)	-0.75 (0.19)	-0.81 (0.06)	0.45 (0.10)	0.30 (0.10)	-0.62 (0.17)	-0.57 (0.13)
Belief Reciprocating, % Participants	89.86 (2.58)	77.42 (7.63)	81.48 (7.62)	64.44 (7.22)	60.00 (8.40)	40.16 (4.37)	91.67 (4.67)	75.76 (7.58)	63.33 (8.95)	53.57 (9.60)

Note. Trusting beliefs standardized using the mean and standard deviation of the first partner at the start of the game. Standard deviations are pooled by first partner appearance, first partner behavior, and second partner appearance.

Table 6. Study 2, Game 2: Regressions of trust-related behavior, trusting beliefs and reciprocating beliefs at the start

| DV:	Sharing Behavior			Trusting beliefs			Belief in Reciprocating		
Model Type:	Logistic			Ordinary Least Squares			Logistic		
Model Number:	(1A)	(1B)	(1C)	(2A)	(2B)	(2C)	(3A)	(3B)	(3C)
P2 Appearance (Trustworthy)	1.11 ** (0.38)	1.11 ** (0.38)	1.14 * (0.44)	1.51 *** (0.15)	1.51 *** (0.15)	1.50 *** (0.18)	1.08 *** (0.29)	1.08 *** (0.29)	1.08 ** (0.34)
P1 Appearance (Trustworthy)	-0.13 (0.35)	-0.03 (0.49)	-0.03 (0.49)	0.04 (0.15)	0.18 (0.21)	0.18 (0.21)	-0.03 (0.28)	-0.35 (0.40)	-0.35 (0.40)
P1 Reciprocated	-0.20 (0.34)	-0.09 (0.52)	-0.09 (0.52)	-0.24 (0.15)	-0.08 (0.22)	-0.08 (0.22)	-0.35 (0.28)	-0.70 (0.42)	-0.70 (0.42)
P1 Appearance x P1 Reciprocated		-0.19 (0.69)	-0.16 (0.73)		-0.30 (0.31)	-0.32 (0.35)		0.65 (0.56)	0.65 (0.61)
P2 Appearance x P1 Trustworthy x P1 Reciprocated			-0.12 (0.85)			0.05 (0.37)			0.00 (0.69)
Constant	1.41 *** (0.34)	1.35 *** (0.39)	1.34 ** (0.40)	-0.88 *** (0.16)	-0.88 *** (0.17)	-0.87 *** (0.18)	0.62 * (0.28)	0.81 * (0.33)	0.81 * (0.34)
Adjusted R^2				0.27	0.27	0.26			
AIC	229.41	231.33	233.31	874.55	875.6	877.58	311.32	311.97	313.97

Note. Numbers in parentheses are standard errors. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Trusting beliefs standardized using the mean and standard deviation of the first partner at the start of the game. Standard deviations are pooled by first partner appearance, first partner behavior, and second partner appearance.
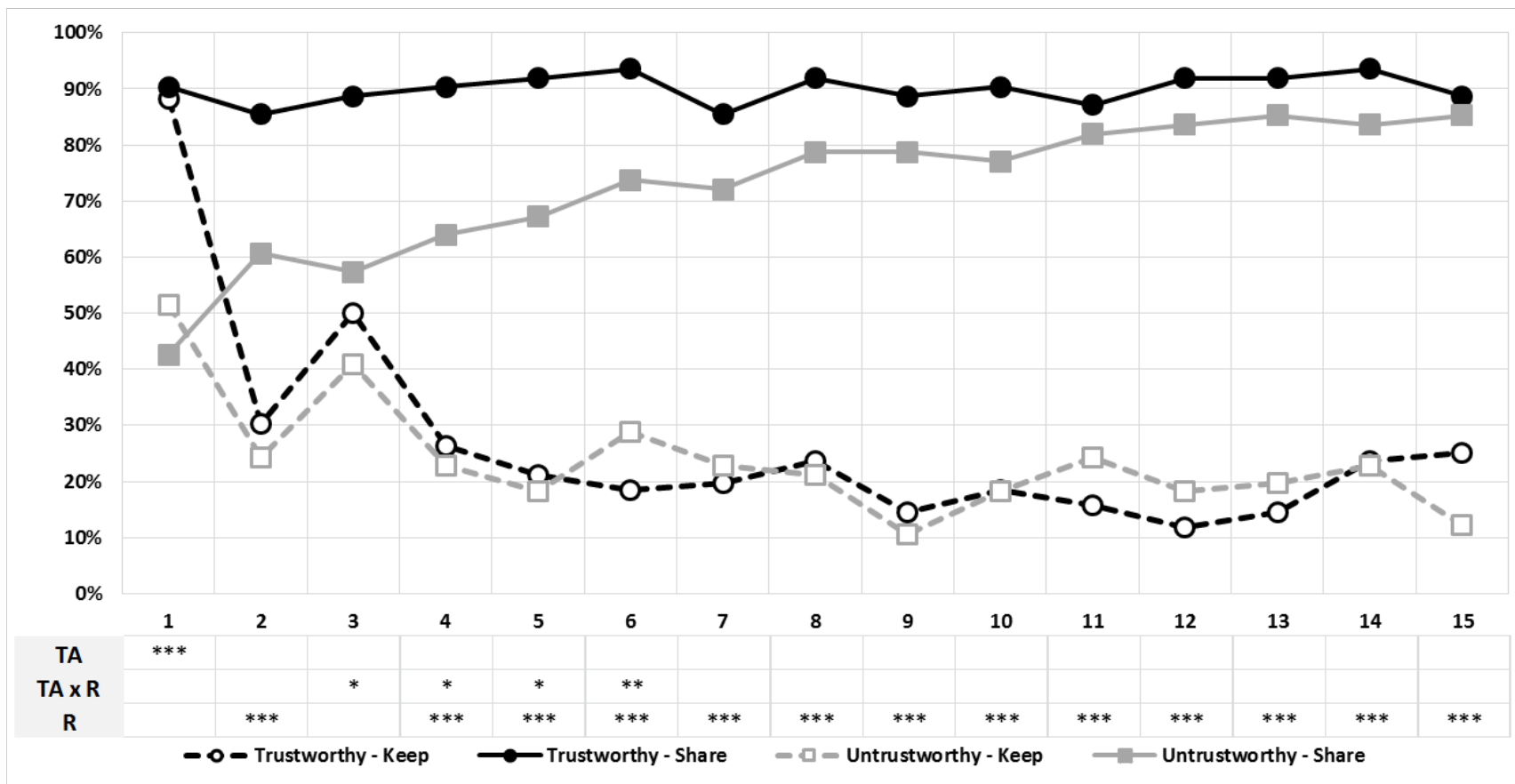
Figure 1. Trust Model
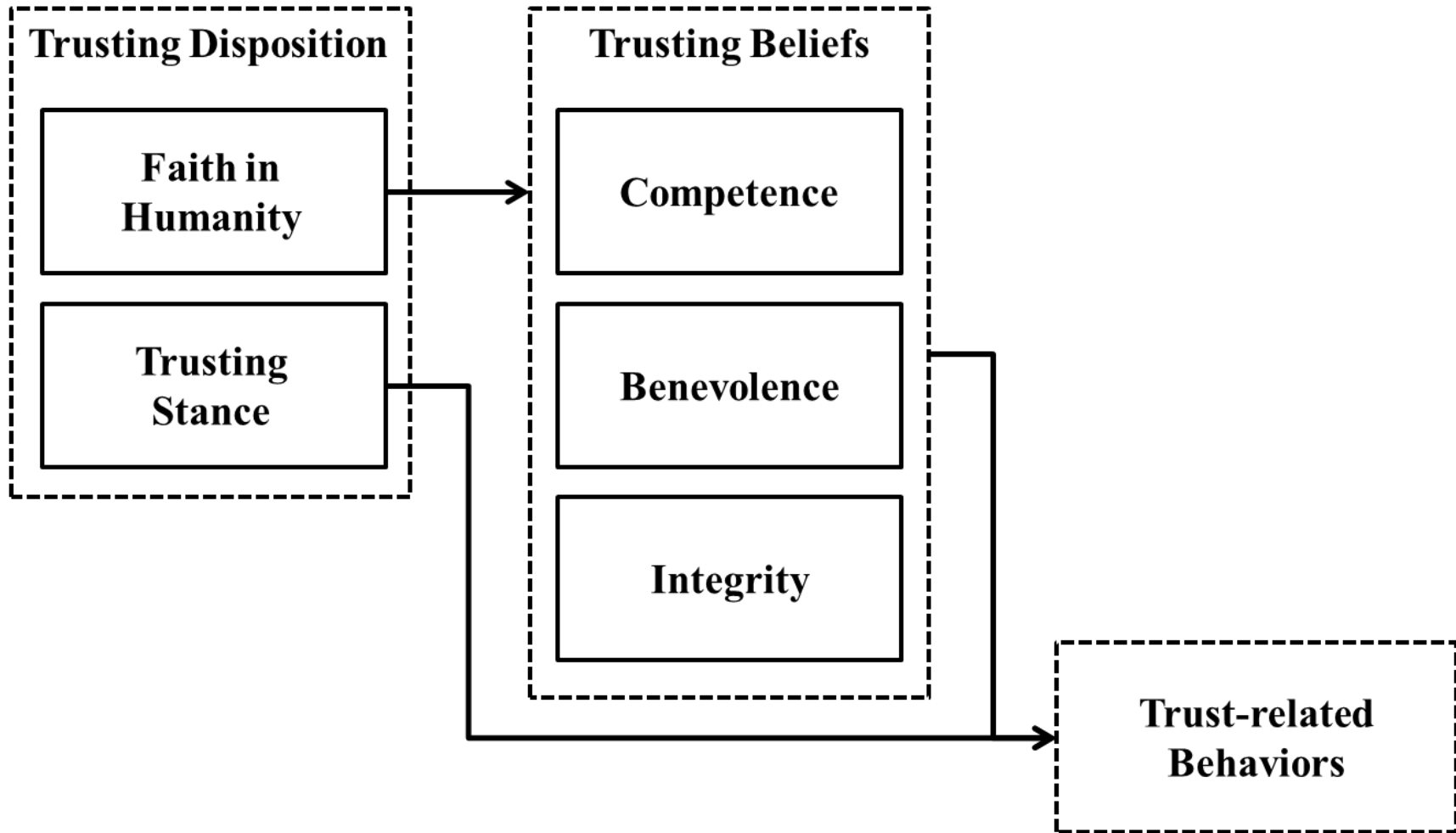
Model adapted from McKnight and Chervany (2002); some components from original model are not included.

Figure 2. Study 1: Percent participants sending points in each round, by treatment
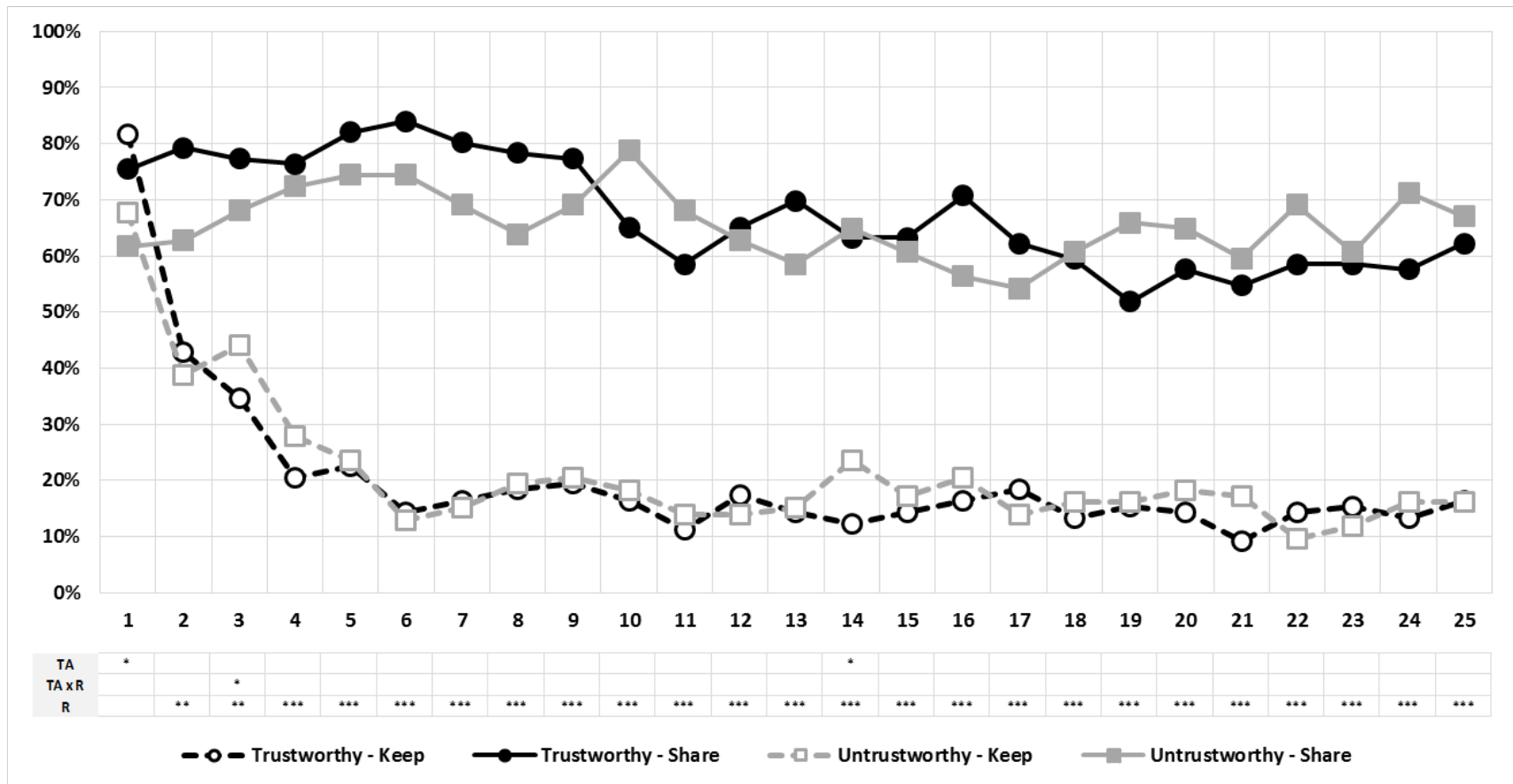
Note: Table at bottom indicates significance level of trustworthy appearance ("TA"), reciprocating behavior ("R"), and the interaction ("TA x R") for the partner in a logistic regression. They are included for illustrative purposes and are not adjusted for multiple comparisons. *** p < 0.001, ** p < 0.01, * p < 0.05

Figure 3. Study 2, Game 1: Percent participants sending points in each round, by treatment

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
TA	***														
TA x R			*	*	*	**									
R		***		***	***	***	***	***	***	***	***	***	***	***	***

- –O– Trustworthy - Keep   —●— Trustworthy - Share   - □ - Untrustworthy - Keep   —■— Untrustworthy - Share

Note: Table at bottom indicates significance level of trustworthy appearance ("TA"), reciprocating behavior ("R"), and the interaction ("TA x R") for the partner in a logistic regression. They are included for illustrative purposes and are not adjusted for multiple comparisons. *** p < 0.001, ** p < 0.01, * p < 0.05