THE UNIVERSITY OF CHICAGO

The Role of Reputation and Perceived Trustworthiness in Cooperation

By

Husamettin Ata Algan

Spring 2024

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the
Master of Arts Program in the Social Sciences

Faculty Advisor: Dr. Alexander Todorov
Preceptor: Dr. Andrew Proctor

Table of Contents

Abstract

How do people evaluate various sources of information to make trust judgments? Specifically, how do people weight the different signals that are presented to them in a cooperative context? Although previous research has investigated the impact of perceived facial trustworthiness and reputation on cooperative behavior, there remains a gap in understanding how reputation is dynamically formed and evolves to influence decisions of trust. In this study (N = 1,778), we explored people's tendency in partner selection and the amount of investment for the chosen partner in a trust game. We explored the influence of the biases in facial trustworthiness on the emergence of reputation (e.g. other's preferences) and its development. Notably, our experiment allowed us to examine these signals in real-time as the reputation information was updated after each decision was made. We also manipulated the salience of reputation with one group viewing faces in descending order of their reputation numbers (the actual choices of prior participants), while another group observing faces presented in a randomized order. We found that when people are presented with male faces, but not with female faces, partner selection is affected by the type of information. In addition, the information type did not affect the investment amount on partners. These findings offer new insights into the dynamic nature of trust judgments in the context of cooperation, and it emphasizes the need for closer examination of the interaction of different informational signals in decision-making.

1) Introduction

1.1 Judgments of Faces

The human face is a unique stimulus that, within milliseconds, activates a variety of trait inferences (Olivola & Todorov, 2010). Among these inferences are social judgments, which include, but are not limited to, evaluations of attractiveness, trustworthiness, competence, and more. Studies have found that quick judgments about traits from faces can have practical implications in the real-world. For example, people can often guess who will win an election just by looking at candidates' faces and judging who seems more competent (Ballew & Todorov, 2007). Also, when judges make quick judgments about how trustworthy criminals look, it can affect their decisions (Porter et al., 2010). And people can distinguish the faces of business, military, and sports leaders better than just by guessing even though they are not able to recognize them (Olivola et al., 2014). Given the significant role that facial judgments play in social interactions and decision-making; it is crucial to explore the underlying mechanisms that drive these perceptions. This leads us to a key part of our social interactions: deciding who we can trust based on their face.

In the absence of direct knowledge, deciding whom to trust is shaped by various biases such as perceived facial trustworthiness (Todorov & Oh, 2021; Olivola et al. 2014). Our brains automatically make inferences about someone's trustworthiness, even though these judgments are inherently biased and far from impartial, highlighting the complex nature of trust evaluations. To put it in another way, when people are faced with ambiguity, they respond based on their personal experiences, history, and knowledge (Albohn et al., 2023). As such, there is no face that everyone can agree on being universally perceived as trustworthy and our judgments from faces are inescapably biased. Nonetheless, there is evidence that judgments are consistent across groups of people (Oosterhof & Todorov, 2008) and shared across cultures (Todorov & Oh, 2021). Research

has shown that the level of trust people place in others can be influenced by the appearance of their faces. Specifically, if a face appears more trustworthy, people are more likely to trust the person it belongs to (van't Wout & Sanfey, 2008). This finding has been explored in experiments in which the facial features of individuals were altered, or 'manipulated', in different ways to make them look more or less trustworthy. For instance, one study showed that people's decision to trust and invest in a person is contingent upon the subjective judgments of facial trustworthiness (Chang et al., 2010). In another study, researchers found that the influence of facial cues is apparent even in children aged 5 to 10 years, where they were more likely to choose partners who appear trustworthy over those who did not (Ewing et al., 2015). Therefore, we have evidence that humans are sensitive to varying levels of trustworthiness, with a tendency to choose faces that seem more trustworthy. This leads us to the second key aspect of our social interactions: deciding whom to trust when we are aware of an individual's actions beyond our personal judgments.

1.2 Social Influence

In the presence of direct knowledge, people have a tendency to rely on informational signals that can shape their behavior. Specifically, because cooperative relationships depend on the foundation of trust, our judgments of trustworthiness are tied to both personal and independent observations, as well as the broader context of social cues and shared experiences. That is, individuals' decisions are shaped by other's actions in addition to their initial judgments (Cialdini & Goldstein, 2004). For instance, as anyone who has ever purchased an item online knows, not every product visual or description reflects the actual quality of the item or features that interest us. Therefore, to get a more accurate or better understanding of the item you are looking at, it can be beneficial to rely on the reviews of the item. Accordingly, our judgments can be affected by intrapersonal and interpersonal dynamics, as we evaluate the available evidence through both

subjective opinion and social feedback. The extent to which such feedback can change a belief depends on how the signal is perceived in terms of its relevance and credibility. If the information is relevant in the context of decision, people can adhere more to the possible benefits of belief change. Likewise, if the information is presented by a source perceived as credible, it is a possibility that beliefs in judgments might change (Bromberg-Martin & Sharot, 2020). Overall, studies on social influence show that decisions are shaped by relevant information about others (Sharot et al., 2023), leading people to alter their existing beliefs based on the information they receive. Most importantly, people with a history of cooperation are preferred in cooperative contexts (Rand & Nowak, 2013).

Social signals can also differ in their salience. For instance, people tend to prefer partners who have shown kind and fair behavior in the past, as evidenced by a study indicating a greater willingness to choose partners based on their historical treatment of others (King-Casas et al., 2005). Another study showed that in financial decisions, individuals are more likely to invest in those perceived as high in moral character, avoiding partners known for venal behavior or even those considered morally neutral (Delgado et al., 2005). Moreover, social signals have the potential to impact the progression of the patterns of behavior. For example, initial advantages in various domains, whether financial, social, or reputational, have been shown to significantly increase the likelihood of subsequent success, demonstrating a universal "success breeds success" phenomenon across different settings (Van de Rijt et al., 2014). Another important study was conducted to test the effect of presenting download counts, which can be considered a form of social signal, on the popularity of songs in an artificially created music market in which the researchers asked participants to listen a given set of songs. They found that the visibility of download counts next to each song impacted song popularity, and even more in the artificial worlds with songs that were

ordered in descending order of download counts (Salganik et al., 2006). In other words, the increased salience of the social signal — through manipulating the ordering of songs — increased the magnification of popularity effects, leading to a more pronounced inequality in song success. However, these results were only partially replicated in a subsequent study conducted by Salganik & Watts (2008). Their experiment followed a similar paradigm with that of the 2006 study but differed in a key aspect: it not only tracked the effect of social signals but also tested the resilience of these effects by inverting the salience of social signal provided to participants. They found that the best songs that were initially manipulated to be the least popular—due to the inversion of download counts—eventually regained their success. In short, the salience of social signals might influence the effects of interest, but the effects of such influences are not strictly deterministic of the outcome.

1.3 Coexistence of initial biases in perceived facial trustworthiness and reputation in cooperation

The literature on facial impressions and social influence has shown that people tend to choose more trustworthy-looking faces, and their decisions can be influenced by others' choices. While most of the research has focused on trust judgments separately, only a small number of studies have explored them together, and these have done so in a static manner, lacking an exploration into the emergence and evolution of the social signals that affect our decision to trust in the context of face judgments.

In a cooperative context, we can argue that reputation can function as a social signal, serving as a critical and valuable source of information that extends beyond the confines of one's immediate social group (Semmann et al., 2005). This information is useful and indicative of the

decisions that are made in trust games, as the literature suggests. In like manner, perceived facial trustworthiness can be informative and useful in deciding who to trust. On one hand, we expect people to choose higher trustworthy looking faces and on the other hand, we expect people's judgments to be influenced by other's choices. If these two components of decision-making—facial trustworthiness and the influence of others' choices—have the potential to dynamically interact in influencing our trust decisions, then it becomes crucial to ask: How does the reputation emerge when people are presented with faces in the context of a cooperative game, and how do initial biases in perceived facial trustworthiness shape its development?

In this study, we investigate whether initial biases in perceived facial trustworthiness can actually shape the emergence of reputation in a trust game. Reputation, in this context, acts as a social signal (e.g., preferences of other players) that accumulates over time and reflects the individual's history of cooperation. However, it is possible that initial biases in perceived trustworthiness influence the emergence of reputation (e.g., more players gravitate toward trustworthy-looking individuals).

Additionally, we will examine the possible bias amplifications that arise from presenting this social signal in people's decision-making processes. How does the introduction of direct knowledge about an individual's past behaviors impact the weight of initial biases in decision-making processes? Do people gravitate towards faces that accumulate higher scores of reputation even though their perception of facial trustworthiness might be in conflict with other's choices? If they do, is this decision more robust in the early stages of the experiment or the weight of social signal doesn't show a clear pattern?

Furthermore, is it possible that the increased salience of reputation information can lead to stronger inclination to choose one face over the other? That is, how does stronger signaling of

reputation influence trust judgements? To investigate how stronger signaling of reputation influences trust judgments, we manipulated the salience of reputation information by presenting faces in descending order of their reputation scores, thereby making the social signal more pronounced than presenting reputation information without any order of hierarchy.

Lastly, how do all of these aspects of trust judgments dynamically evolve and interact over time? To this end, no study has observed the role of perceived facial trustworthiness and reputation in real time, making this study an important step to observe how the emergence and evolution of reputation is influenced by perceived facial trustworthiness.

In summary, the present study aims to answer the research question: “How does the dynamic updating of reputation knowledge socially influence an individual's choices in cooperative contexts?”. In light of previous work on face impressions and social influence, we hypothesize that:

H1: There will be a higher likelihood of choosing trustworthy-looking faces.

H2: Signaling participants with reputation scores will affect partner selection. Increasing the salience of reputation by presenting faces in descending order of these scores should increase the likelihood of choosing those that appear trustworthy.

H3: More trustworthy-looking faces will accumulate reputation scores faster and that this would in turn increase the likelihood of choosing them.

2) Methods

Participants:

A sample of 1,800 participants was recruited through CloudResearch's MTurk Toolkit, a platform commonly used for online data collection, with the recruitment process being conducted in batches of 100. Of the 1,800 participants, 1,778 were included in the analysis for two reasons: 1) 10 participants did not fully complete the study, as evidenced by their responses to the open-ended questions at the end of the experiment, and 2) 1 participant appeared twice in the data set. In addition, as indicated in the preregistration form for the study, we excluded participants that finished the experiment in less than 1.5 minutes. This was done to prevent the inclusion of data from individuals who may not have engaged with the material thoroughly, considering that the expected duration of the experiment was around 5 minutes. As a result, the final sample included 1,778 participants (M_{age}= 42.57, SD_{age}= 12.33) divided into six conditions: Baseline Male, Baseline Female, Random Male, Random Female, Ordered Male, and Ordered Female. The Baseline Male condition included 299 participants, Baseline Female included 296 participants, Random Male included 297 participants, Random Female included 294 participants, Ordered Male included 297 participants, and Ordered Female included 295 participants.

Participants self-identified as follows: 942 male, 790 female, with an additional 9 identifying in other categories and 13 preferring not to disclose their gender; 1302 White, 162 Black/African American, 97 Mixed, 78 East Asian, 63 Latinx/a/o or Hispanic, 30 Southeast Asian, 19 South Asian, 12 Other/Not reported, 9 American Indian or Native American, 3 Middle Eastern, 3 Race/Ethnicity not listed.

Stimuli:

The experiment consisted of 6 male and female faces. Importantly, the 6 hyper-realistic faces were generated by a data-driven computational model of perceived trustworthiness (Peterson et al., 2022). These 6 faces correspond to different levels of perceived trustworthiness from 20 to 80 in increments of 12 (20, 32, 44, 56, 68, 80) out of 100. Below, Figures 1 and 2 show the face stimuli with their perceived trust levels.

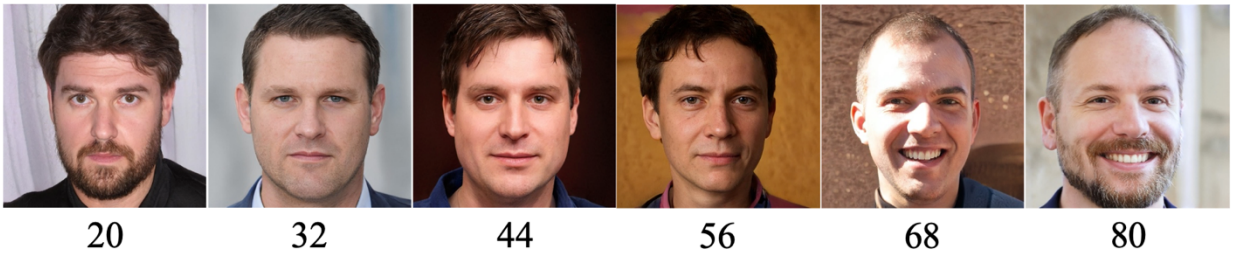

Figure 1: Male faces with their perceived trustworthiness scores.

Figure 2: Female faces with their perceived trustworthiness scores.

Experiment Procedure:

After giving their consent, participants were directed to a screener website that was intended to discourage bots from joining. Then, the screener page asked participants to input "I will answer open-ended questions" exactly as printed into a text box. After passing the screener, participants were told that they would be playing a simple game in which they would choose a partner from a list of potential partners. Participants were also instructed that they had been given a gift of $1, which they could choose to share with a partner.

They could offer any amount from $0.00 to the full $1.00. If shared, the amount would be quadrupled when given to the partner. Subsequently, the partner could decide to keep the entire amount or share it back, with the potential for the participant to receive double what they initially gave. Additionally, participants were told to select a partner by choosing from one of six faces displayed in a grid. They were informed that their counterpart's decision—whether to cooperate or keep the gift—might have been made previously, allowing immediate disclosure of the outcome. Finally, after the summary slide of the experimental procedure, participants were presented with the faces corresponding to the condition group to which they had been assigned.

The experiment's conditions varied based on the presence or absence of social signals, the salience of these social signals, and the gender of the facial stimuli. Conditions were categorized as 'baseline', where participants made choices without knowledge of prior participants' decisions, meaning they had no access to the reputational information about the partners presented to them. In the study, participants were shown the choices of prior participants under what were termed "reputation conditions." Within the ‘reputation’ conditions, there were two sub-categories: 'random' and 'ordered.' These categories differed in how the facial stimuli were presented. Specifically, the salience of the signal was manipulated by changing the order in which these stimuli appeared in each category. In other words, the salience of reputation knowledge was manipulated with one group viewing faces in descending order of their reputation numbers (the actual choices of prior participants) while another group observing faces presented in a randomized order. But how do the participants assigned to the reputation conditions see the reputation score of each face? On top of each face, there is a dark blue colored rectangle with numbers on it that was intended to show how many times each person was chosen as a partner by the previous players. Initially, the first participants to ever start the experiment in both the random and ordered

conditions would see zeroes, indicating that at the start, previous players had made no selections in the game. In other words, no partner was selected as a partner in the trust game. For clarity, the following figures will present these details for the male stimuli, but it can be assumed that the same applies to the female stimuli as well. Therefore, the first participant assigned to any of the two reputation conditions with male stimuli would see an array of faces like in Figure 3.

Figure 3: Array of faces presented to the participant in 'Random Male' and 'Ordered Male' conditions at t = 0.

Let's further elaborate on the process of updating these scores and how it affects the display of faces. For example, in the 'random male' condition, if the first participant chooses the rightmost face as their partner, the next participant would see that the reputation score of this chosen face has 1 on top of it. However, the sequence in which faces are displayed to this second participant in the random condition remains random, not influenced by the updated scores.

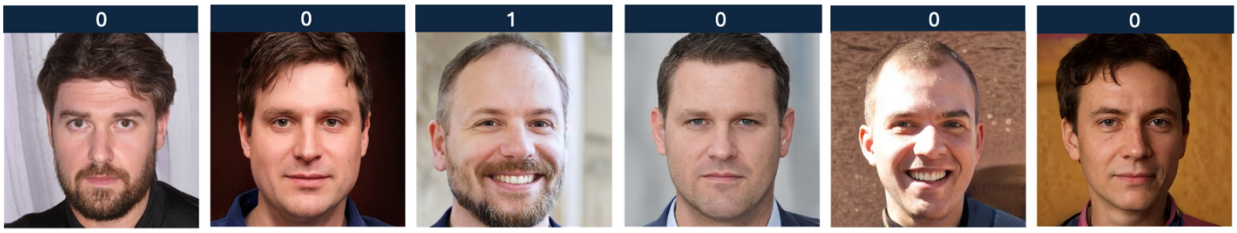

Figure 4: Array of faces presented to the participant in 'Random Male' condition at t = 1.

In the 'ordered' conditions, however, the ordering of faces is different than the 'random' conditions. Following the example, if a participant chooses the same face, the subsequent participant will see the faces arranged not randomly but in a sequence that reflects their reputation

scores—faces with higher scores are displayed in descending order of reputation scores. Accordingly, the second participant in the ordered male condition would see the following array of faces:

Figure 5: Array of faces presented to the participant in 'Ordered Male' condition at t = 1.

After establishing how participants in the reputation conditions are presented with the choices of prior participants, it's crucial to understand the dynamics of how these reputation scores are updated in real-time. With each completed trial, the experiment updates the reputation scores to reflect the latest prior choices. It's important to note that only the first participants in these reputation conditions encounter a '0' atop each face, signifying that up until their participation, no selections have been made, leaving all potential partners without an accumulated reputation score.

As participants make choices, reputation scores are adjusted to reflect the cumulative decisions of all preceding participants. Essentially, our experimental setup allows us to gather and update reputation data in real time. This means that with each new participant, the reputation scores displayed on faces are revised to reflect the previous participants' choices. In addition, the ordering of faces is always randomized in the baseline and random conditions whereas the array of faces in ordered condition is in descending order of their reputation scores. Below, you can see the comparison of the last trial for random and ordered male conditions.

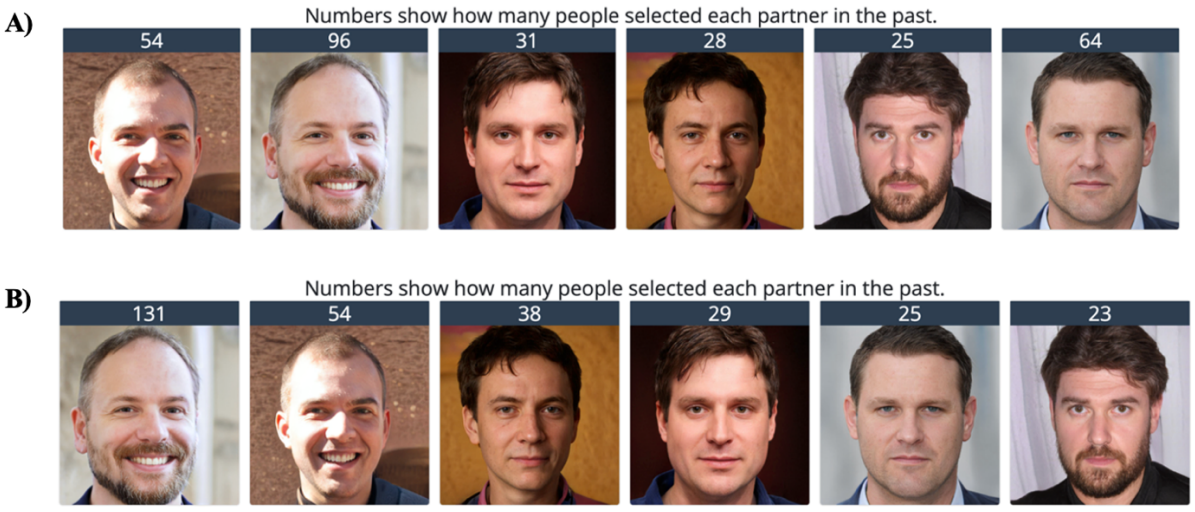

Figure 6: Male faces displayed in random order of reputation numbers (in A, 'random male') and male faces displayed in descending order of reputation numbers (in B, 'ordered male') in the last trial of the respective conditions.

After making their choice, they were asked to indicate the amount of their endowment they would invest in the chosen partner on a scale from $0 to $1 using a simple slider.

Lastly, participants filled out the open-ended questions that consisted of debriefing questions (“In one or two sentences, did you use any particular strategies when choosing your partner?”, “Did you interrupt your completion of the survey in any way (e.g. by switching to other windows or tabs on your computer, or by refreshing the task page)?”, “Have you ever participated in this study before?”, “Was any part of the procedure unclear? Did you have any problems completing the task?”, “We know it is generally difficult to stay focused in these online experiments, especially when they take this long. On a scale of 1-100 (with 1 being very distracted, and 100 being very focused), how well did you pay attention to the experiment?”, “Did you notice anything unusual about the faces you saw? If so, what?”, “How much did you believe that you were playing with other real people while playing the game? Please answer on a scale of 0-100 (with 0 meaning "Did not believe at all", and 100 meaning "Believed completely").”, “Do you have any other comments you would like to add? Feel free to tell us whatever you thought about

the task, or any problems you might have encountered (that you didn't already mention above).") followed by a short demographics questionnaire (age, race, and gender).

3) Results

The results from this study aimed to test hypotheses related to the dynamics of reputation knowledge and its impact on cooperative choices with regards to how individuals select partners for cooperation. Firstly, we anticipated a higher likelihood of participants choosing faces that appeared more trustworthy from an array of options. Secondly, we hypothesized that the accumulation of reputation information—specifically, the choices made by previous participants—would affect the selection of faces, particularly when these faces were presented in an order of descending reputation scores. Lastly, we expected that faces perceived as more trustworthy would garner reputation scores more quickly, thereby increasing their chances of being chosen by subsequent participants.

In addition, people's willingness to share the amount they received as a gift with their chosen partner was analyzed as the second key dependent variable in the experiment. Specifically, the effect of perceived trustworthiness of the chosen face, the presentation of the reputation knowledge, the reputation number of the chosen face, and the interaction of all of these with respect to shared amount.

It should be noted that across all visualizations and statistical analysis, we employed LOESS (locally estimated scatterplot smoothing). This was done due to the fact that, during the data collection, some number of participants saw the exact array of reputation numbers. For that reason, the LOESS method provided the smoothing of data points without imposing any predetermined distribution shape on the dataset. Additionally, we could better represent the

subtleties of our dataset and the true trends without the limitations of parametric assumptions by using a non-parametric smoother.

3.1) Partner Selection

3.1A. Do people choose higher trustworthy-looking faces?

Firstly, to better understand if participants were more likely to choose faces that appeared trustworthy, we created a bar graph that shows the overall percentage of times each face was selected. In that regard, we chose to plot a bar graph for the aggregate proportion of choosing each face, irrespective of the condition or gender group they were assigned to, which resulted in Graph 1.

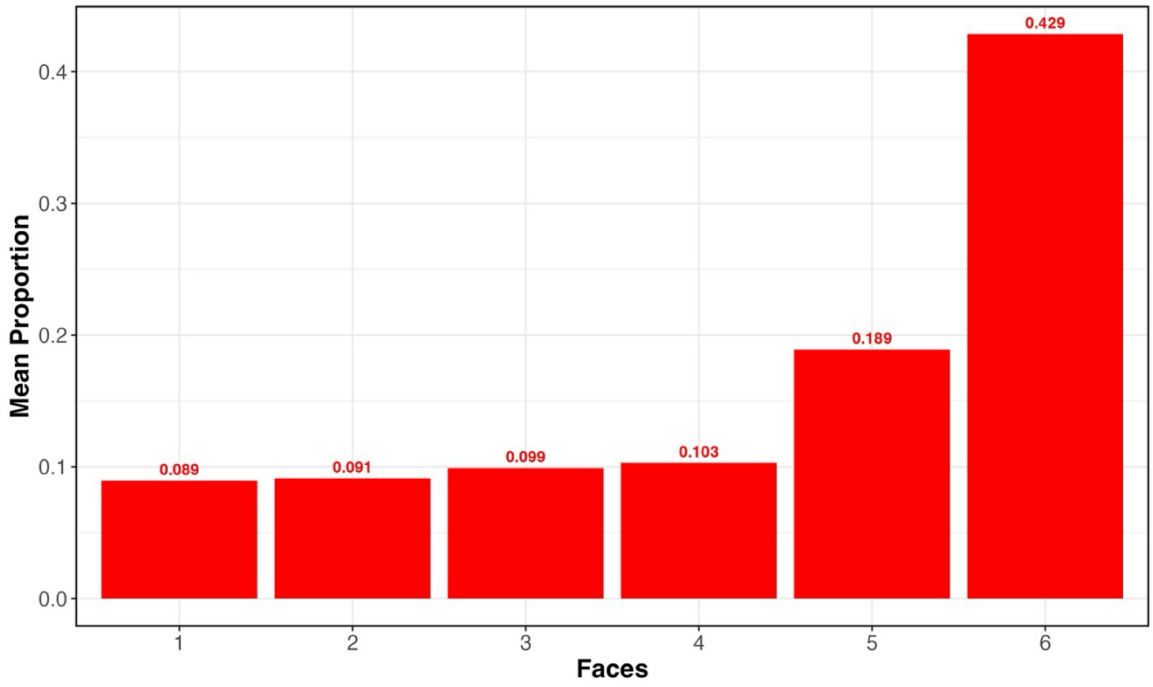

Figure 7: Mean proportion of choosing each face.

We can see that as perceived trustworthiness level increased, participants showed a higher likelihood of selecting those faces as potential partners. Face 1, which had the lowest perceived trustworthiness had a mean selection frequency value of about 0.089. This was followed by Face

2 with a value of 0.091, Face 3 with 0.099, Face 4 with 0.103, Face 5 with 0.189, and Face 6 with the highest value of 0.429. As shown in Graph 1, there is a notable difference between mean proportion of Faces 1 to 4, with perceived trust levels of 20, 32, 44, and 56, respectively, and Faces 5 and 6, which have trust levels of 68 and 80, respectively.

However, Graph 1 does not account for the critical subgroup variations relevant to the research question. Consequently, a more detailed analysis was conducted, focusing on the preferences for partner selection within each condition group. Below, Graph 2 depicts the mean proportion of choosing each of six faces across three different condition groups (Info): Baseline, Random, and Ordered. Each condition group is represented by a different color—Baseline (black), Random (green), and Ordered (orange).

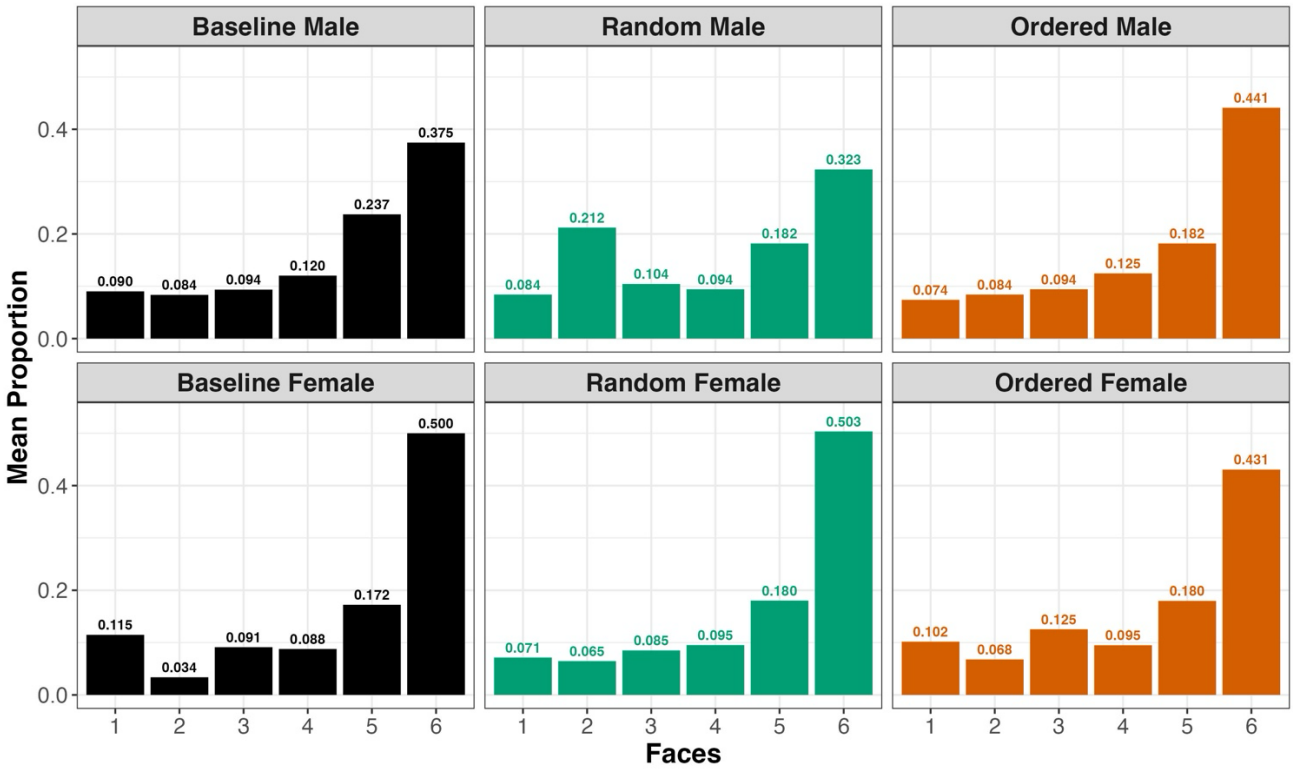

Figure 8: Mean proportion of choosing each face across conditions.

In the baseline conditions, the face associated with a lower trust level (Face 2) was the least preferred among both genders, whereas the face associated with the highest trust level (Face 6)

was favored the most, especially in female stimuli. In the random conditions, preferences notably diverged following the introduction of reputation scores. Among male stimuli, when faces were randomly presented with reputation scores, the face with a lower trust level (Face 2) outperformed the faces with a higher trust level (Faces 3, 4, and 5). This suggests that sharing information about reputation influenced participants' choices towards male faces, favoring those with specific trust levels over others by the end of the experiment. For female faces, when reputation information was made available, a clear and consistent preference can be observed, with faces having higher perceived trustworthiness being selected more frequently than lower trustworthy faces. Furthermore, when comparing the baseline and random conditions, distinct gender-specific trends become apparent, and the introduction of reputation information appear to magnify the existing biases, as observed by the differences in proportions of faces with the same levels of perceived trustworthiness. In the ordered conditions, where faces were displayed in order from highest to lowest reputation scores, the effect on partner selection was most pronounced for the highest trust level face (Face 6) among conditions with male stimuli. Conversely, the likelihood of selecting this face among female stimuli was reduced compared to both the random and baseline conditions.

To analyze the patterns shown in Figure 2 and assess the statistical significance of differences in face selection across experimental conditions and levels of perceived trustworthiness, Pearson's Chi-squared tests were performed. These tests were designed to assess whether the frequencies of face selections were independent of both the experimental conditions and the perceived trustworthiness of each face. Results showed that the relationship between perceived trustworthiness and information types was significant for male faces, X^2 (10, N = 964) = 36.41, p <.001 and insignificant for female faces, X^2 (10, N = 790) = 11.85, p = 0.2949. These results suggest that, while there is a clear pattern of preference for more trustworthy-looking faces

among the male faces as a function of experimental condition, this pattern does not hold for the female faces. Conversely, the lack of a statistically significant difference in the selection of female faces across varying experimental conditions suggests that participants' preferences for female faces were not influenced by the perceived trustworthiness levels presented. This indicates a consistent selection behavior towards female faces, regardless of the information provided about their trustworthiness in the experimental setup.

In summary, participants assigned to the conditions with male faces were significantly influenced by the information in which the faces were presented. This implies that the male faces' perceived trustworthiness may be more susceptible to the effects of situational variables such as the order in which information is presented or the presence of additional contextual details. For female faces, however, the data suggests a different trend. The lack of significant results across conditions points to a baseline preference that is robust against such experimental manipulations.

3.1B. How does the presence of reputation influence trust?

Expanding on the first hypothesis, the next inquiry in our research question tests whether the presence and the salience of reputation knowledge — that is, the order in which partner choices are presented to the participants — influences cooperative behavior. Graphs 1 and 2 did not account for an exploration into the effect of accumulating reputation scores on people’s partner choice. Therefore, the purpose of Graph 3 was to provide a visual representation of the relationship between a face's reputation score and the frequency with which it is chosen in conditions with reputation knowledge. Hence, Graph 3 shows how the frequency of face selection varies with accumulating reputation scores across different types of information.

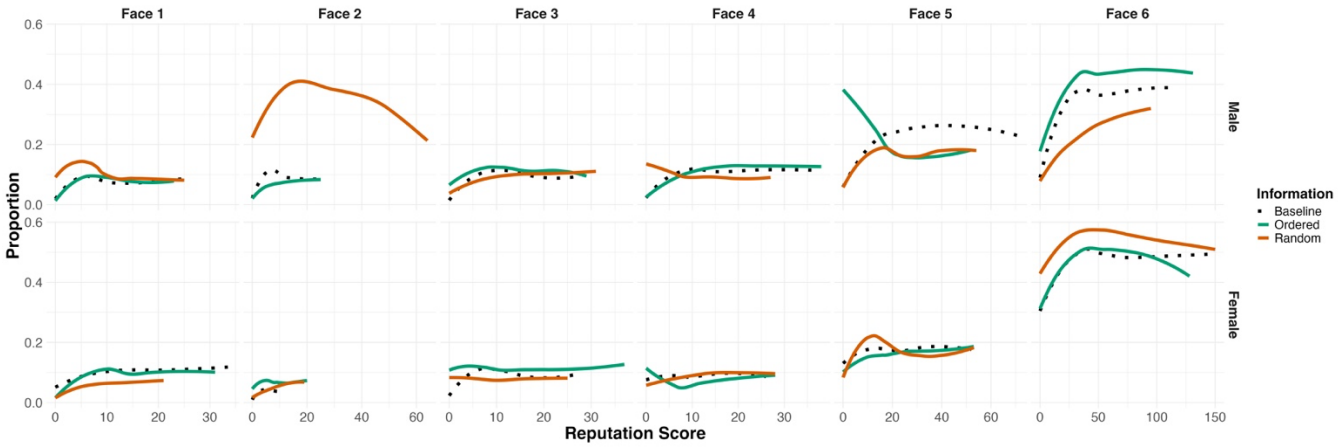

Figure 9: Proportion of Selected Faces by Accumulating Reputation Scores Across Information Types.

In order to see how the presentation and accumulation of reputation scores affect participants' choices of faces, this relationship was investigated for each face. Moreover, the baseline trend was computed and shown to allow for comparison between conditions with the absence and presence of reputation scores.

We observed that the information type lines follow a close pattern to each other for faces 1, 3, and 4. However, the male face 2 in the random condition was chosen more frequently up until the 20th trial, after which its selection rate gradually decreased. Nonetheless, it stayed as a winner compared to faces with lower perceived trustworthiness.

There were clearer variations in the trends of proportion as the reputation scores increased, especially visible in faces 5 and 6 for both gender stimuli. Notably, there was a significant drop in the selection rate of face 5 compared to all other conditions and faces among the male stimuli. Furthermore, while presenting the faces in a specific order increased the selection of face 6, which had the highest score of perceived trustworthiness, random presentation did not have the same effect on the selection of male faces. On the other hand, randomly presenting face 6 increased the frequency with which this face was chosen, while ordering the faces did not affect its selection compared to the baseline condition among female stimuli. It's also important to mention that

arranging the faces by reputation actually decreased the frequency of selecting face 6 after the 100th trial.

Considering these observations from Graph 3, it is challenging to draw direct conclusions about participants' decision-making processes. For that reason, to examine hypothesis H2 statistically, we conducted a log-linear analysis with separate generalized linear models (GLMs) for male and female faces, using a Poisson distribution with a log link function. These models incorporated the interaction between condition and trust level as predictors, allowing us to examine how these factors jointly influence the count of cooperative actions within each gender group.

	Estimate	Std. Error	Z-Value	P-Value
(Intercept)	**3.2958**	**0.1925**	**17.126**	**< 2e-16*** **
Ordered	-0.2048	0.2872	-0.713	0.476
Random	-0.0770	0.2776	-0.277	0.782
Trust Level = 32	-0.0770	0.2776	-0.277	0.782
Trust Level = 44	0.0364	0.2697	0.135	0.893
Trust Level = 56	0.2877	0.2546	1.130	0.259
Trust Level = 68	**0.9668**	**0.2261**	**4.276**	**< 2e-16*** **
Trust Level = 80	**1.4227**	**0.2144**	**6.636**	**< 2e-16*** **
Ordered * 32	0.2048	0.4031	0.508	0.611
Random * 32	**1.0012**	**0.3646**	**2.746**	**0.006** **
Ordered * 44	0.2048	0.3923	0.522	0.602
Random * 44	0.1787	0.3808	0.469	0.639
Ordered * 56	0.2322	0.3705	0.627	0.531
Random * 56	-0.1744	0.3749	-0.465	0.642
Ordered * 68	-0.0689	0.3393	-0.203	0.839
Random * 68	-0.1967	0.3311	-0.594	0.552
Ordered * 80	0.3615	0.3147	1.149	0.251
Random * 80	-0.0772	0.3105	-0.249	0.804

Table 1: Log-linear analysis summary for male faces on information (baseline, random, ordered) and face trustworthiness.
*Notes: * p <.05, ** p <.01, *** p <.001.*

Table 1 represents the table for log-linear analysis for male faces on information and face trust levels. Contrary to initial expectations, the ordered presentation of information does not have a statistically significant main effect on the likelihood of partner selection, as evidenced by a p-value of 0.476, suggesting that an ordered presentation of faces does not significantly alter the chances of being chosen as a partner compared to a baseline condition. However, the likelihood of choosing a face is positively correlated with its perceived trustworthiness at higher levels. Specifically, trust levels 68 and 80 show a highly significant positive association (p < 2e-16, p < 2e-16) with the choice of the face, with an estimated increase in log odds of 0.9668 for trust level 68 and even higher increase in log odds of 1.4227 for face with trust level 80.

In terms of presenting faces in a random sequence in relation to their reputation scores, we observe that the condition of randomness itself does not significantly affect the likelihood of face selection. On the other hand, there is a significant (p = 006) interaction between the random presentation and the face with a perceived trustworthiness level of 32. This indicates that when faces are presented randomly, those with a trustworthiness level of 32 are more likely to be chosen compared to other levels, with an estimated increase in log odds of 1.0012.

	Estimate	Std. Error	Z-Value	P-Value
(Intercept)	**3.5264**	**0.1715**	**20.562**	**< 2e-16*** **
Ordered	-0.1252	0.2505	-0.500	0.617
Random	-0.4818	0.2775	-1.736	0.083
Trust Level = 32	**-1.2238**	**0.3597**	**-3.402**	**<.001*** **
Trust Level = 44	-0.2305	0.2578	-0.894	0.371
Trust Level = 56	-0.2683	0.2605	-1.030	0.303
Trust Level = 68	0.4055	0.2214	1.831	0.067

Trust Level = 80	**1.4709**	**0.1902**	**7.734**	**< 2e-16*** **
Ordered * 32	0.8183	0.4612	1.774	0.076
Random * 32	**1.1237**	**0.4792**	**2.345**	**0.019***
Ordered * 44	0.4402	0.3561	1.236	0.216
Random * 44	0.4049	0.3925	1.032	0.302
Ordered * 56	0.1993	0.3700	0.538	0.590
Random * 56	0.5559	0.3889	1.430	0.153
Ordered * 68	0.1636	0.3182	0.514	0.607
Random * 68	0.5203	0.3399	1.531	0.126
Ordered * 80	-0.0279	0.2782	-0.100	0.920
Random * 80	0.4818	0.3009	1.601	0.109

Table 2: Log-linear analysis summary for female faces on information (baseline, random, ordered) and face trustworthiness.
*Notes: * p <.05, ** p <.01, *** p <.001.*

Table 2 represents the table for log-linear analysis for female faces on information and face trust levels. For the face that had the highest perceived trust level, there was a highly significant positive association with the choice of the face (p < 2e-16). Similar to male faces, the ordered presentation of reputation scores itself doesn't affect the partner choices, as indicated by an estimate of -0.1252 and a p-value of 0.6173. Moreover, even when the ordered presentation was combined with different trust levels, the interactions did not reach statistical significance across all levels of perceived trust, indicating that the ordered sequence of faces, regardless of their trustworthiness level, does not substantially alter selection probabilities.

In terms of the random presentation of faces, the presentation itself doesn't significantly affect the partner choice, as indicated by an estimate of -0.4818 and a p-value of 0.0826. However, the interaction between random presentation and the face with a trust level of 32 yields a significant positive estimate of 1.1237 with a p-value of 0.0190, which indicates that the combination of random presentation and a trust level of 32 significantly increases the likelihood of selection.

Overall, the analysis conducted through log-linear models for both male and female faces indicate that compared to the baseline information, where no reputation scores were shown, neither random nor ordered presentation of reputation scores significantly altered partner selection probabilities, suggesting that neither presence of reputation knowledge nor the sequence in which faces were presented with reputation scores played a decisive role in cooperative decision-making processes. However, there was a significant association between higher trust levels and partner choice. Interestingly, for both genders, there was a significant interaction of random presentation and face with trust level 32, which might suggest that under certain conditions, the randomness in presentation can subtly influence preferences.

3.1C. Does the presentation of reputation affect partner choice?

As an extension of H2, H3 was tested with conducting an ordinal regression separately for the gender of the face stimuli. In both regression analyses, the dependent variable was the perceived trustworthiness level assigned to the faces chosen by participants. The independent variables included a categorical variable that differentiated whether the faces were presented in an ordered sequence or at random (to capture the effect of presentation order on perceived trustworthiness, referred to here as "reputation"). Additionally, we included the reputation score that was associated with each participant's chosen face. This score was treated as a covariate to accommodate its continuous nature and to examine its linear relationship with the perceived trustworthiness of the faces. We also explored the interaction effect between the type of face presentation (ordered vs. random) and the participant's cumulative reputation score, to understand how these factors jointly influence trustworthiness perceptions.

Coefficients:				
	Value	Standard Error	T-Value	P-Value
Ordered Presentation	-0.1689	0.2543	-0.66	0.506
Reputation Score	**0.0646**	**0.0054**	**7.82**	**<.001**
Ordered Rep. * Rep Score	0.0225	0.0084	2.70	0.007
Intercepts:				
	Value	Standard Error	T-Value	P-Value
20|32	-1.3946	0.2082	-6.70	<.001
32|44	-0.0729	0.1772	-0.41	0.681
44|56	0.5001	0.1754	2.85	0.004
56|68	1.0956	0.1800	6.09	<.001
68|80	2.1548	0.1999	10.78	<.001

Table 3: Ordinal logistic regression summary for male faces

Table 3 shows the ordinal logistic regression summary for male faces, highlighting the effects of presentation type and reputation score on the perceived trustworthiness level. The baseline for comparison is the random condition, wherein faces are displayed in a random order. Moving into the analysis of these results, we can see that holding reputation score constant, the ordered presentation of faces is associated with a slight decrease in the likelihood of choosing a face with a higher trustworthiness level for male faces, although the effect is not strong (t = -0.66, p = 0.506). Moreover, the coefficient value of 0.0646 indicates that a higher reputation score is associated with an increased likelihood of choosing a face with a higher trustworthiness level for male faces, and this effect is statistically significant (t = 7.82, p < .001). Finally, the interaction between the ordered representation of faces and reputation score tells us that the effect of reputation score on the likelihood of choosing a higher trustworthiness level face is slightly more

pronounced in the ordered presentation condition for male faces, which is also significant (t = 2.70, p = 0.007).

In summary, the coefficients suggest that while ordered presentation slightly decreases the likelihood of choosing higher trustworthiness levels, reputation score significantly increases it, especially in an ordered presentation.

Coefficients:				
	Value	Standard Error	T-Value	P-Value
Ordered Presentation	-0.0899	0.2683	-0.34	0.738
Reputation Score	**0.0737**	**0.0077**	**9.63**	**<.001**
Ordered Rep. * Rep Score	-0.0075	0.0098	-0.76	0.446
Intercepts:				
	Value	Standard Error	T-Value	P-Value
20|32	-1.0043	0.2212	-4.54	<.001
32|44	-0.2721	0.2042	-1.33	0.183
44|56	0.5584	0.1999	2.79	0.005
56|68	1.2144	0.2048	5.93	<.001
68|80	2.4834	0.2295	10.82	<.001

Table 4: Ordinal logistic regression summary for female faces

Table 4 shows the ordinal logistic regression summary for female faces, highlighting the effects of presentation type and reputation score on the perceived trustworthiness level. Similar to Table 3, the baseline for comparison is the random condition. Moving into the analysis of these results, there is a slight, though not statistically significant (t = -0.34, p = 0.738), decrease in the likelihood of choosing a face with a higher trustworthiness level for female faces represented in descending order of reputation scores. Conversely, the reputation score coefficient of 0.0737

indicates a robust and statistically significant (t = 9.63, p < .001) positive association with the perceived trustworthiness of female faces, implying that an increase in the reputation score of a chosen face increases its perceived trustworthiness. Lastly, the interaction between the ordered representation of faces and reputation score tells us that a slight, but not significant (t = -0.76, p = 0.446), decrease in the impact of reputation score on trustworthiness ratings within the ordered presentation context, suggesting that the order of presentation may slightly moderate the positive effect of reputation scores on perceived trustworthiness.

In summary, while the ordered presentation in female faces slightly reduces the likelihood of higher trustworthiness ratings, this effect doesn't reach statistical significance. Conversely, the clear and significant positive impact of reputation scores underscores their importance in shaping trustworthiness perceptions. The interaction analysis further suggests that the presentation order may subtly affect the strength of this relationship, though not significantly.

3.2) Sharing Behavior

3.2A. What is the propensity to share in relation to perceived trustworthiness?

As stated in the methods section of this thesis, each participant was asked to indicate their willingness to share with their chosen partner, allowing the analysis of decision-making in this context to specifically explore how initial judgments based on facial trustworthiness and accumulated reputation information influence a person's decision to share their money with others in the game.

Overall, participants invested a relatively high amount of money with their endowment of $1. In addition, the invested amount was similar between the baseline and reputation conditions. Figure 10 presents the mean invested amount for each participant’s chosen partner.

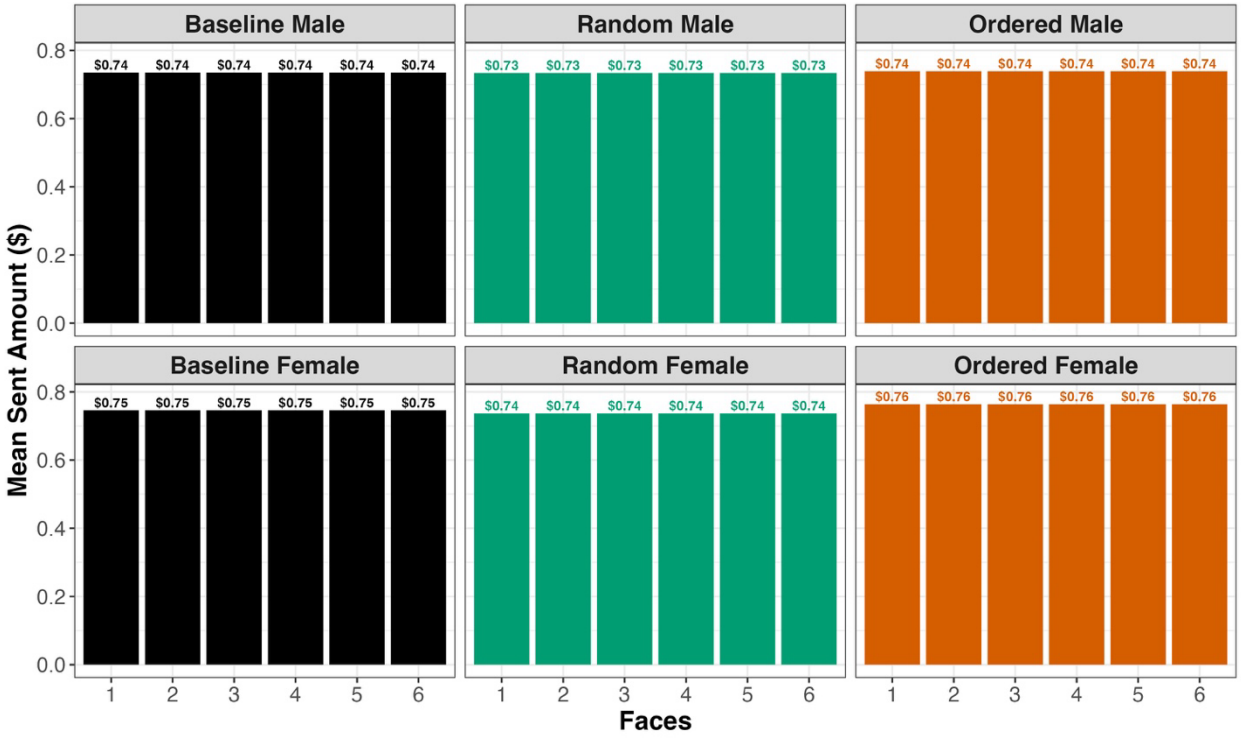

Figure 10: Mean Proportion of Sent Amount Across Conditions

As we can see, the invested amount for each face was exactly the same within each condition and in the absence of reputation scores, participants invested $0.74 cents for the male stimuli and $0.75 for the female stimuli. The addition of reputation numbers changed the baseline invested amounts $0.73 for male stimuli and $0.74 for female stimuli. In addition, when the reputation numbers were salient, the invested amounts changed to $0.74 for male stimuli and $0.76 for female stimuli. In short, there was relatively high level if investment to chosen partners and the variation between different levels of perceived trustworthiness was constant and minimally different within conditions.

To statistically test this second key dependent variable, linear regressions were conducted in two ways: 1) looking at how much money participants chose to invest based on how trustworthy they thought their chosen faces were, which experimental group they were in, whether the faces were male or female, and how all these factors worked together; and 2) focusing on the groups that

had reputation information, to see how the trustworthiness of the faces, the type of reputation information (random or ordered), the gender of the faces, and their reputation scores influenced the amount of money invested, including how all these factors interacted.

Table 5 shows the summary of the linear regression for the amount sent, based on the trustworthiness level of the participant's chosen partner and the condition, for male stimuli.

Coefficients:				
	Estimate	**Standard Deviation**	**T-Value**	**Pr(>|t|)**
(Intercept)	**0.6319**	**0.0415**	**15.22**	**<2e-16*** **
Trust Level = 32	0.0187	0.0500	0.37	0.7088
Trust Level = 44	**0.1045**	**0.0526**	**1.99**	**0.0473***
Trust Level = 56	0.0948	0.0509	1.86	0.0629
Trust Level = 68	**0.1163**	**0.0460**	**2.53**	**0.0116***
Trust Level = 80	**0.1413**	**0.0427**	**3.31**	**0.0010*** **
Ordered	0.0004	0.0273	0.01	0.9891
Random	0.0117	0.0276	0.42	0.6729

Table 5: Linear regression analysis summary for sent amount based on trustworthiness level and condition in male stimuli.
*Notes: * p <.05, ** p <.01, *** p <.001.*

The intercept in the regression analysis represents the average amount sent when the trust level is at the reference point of 20, which is the lowest level considered, and when no reputation information is presented (the 'baseline' condition). This intercept amount is $0.6319 and is significantly different from zero ($SD = 0.0415$, $t = 15.22$, $p < 0.0001$), suggesting that participants tend to send a notable amount of money even when they have no information about the partner's trustworthiness or reputation.

As trust levels rise above the reference point to 44, 68, and 80, the amounts sent increase significantly, with respective p-values of 0.0473, 0.0116, and 0.0010. This indicates that participants are likely to send more money as the perceived trustworthiness of a face increases

from the reference level of 20. When the perceived trustworthiness is at levels 32 and 56, the increments in the amount sent compared to the reference level do not show statistical significance. At trust level 32, the p-value is 0.7088, and at trust level 56, it's 0.0629. This implies that perceived trustworthiness at these levels does not significantly alter the sending behavior compared to the reference trust level.

Regarding the experimental conditions, neither the 'Ordered' nor the 'Random' presentation of faces affects the amount sent in any significant way (p = 0.9891 for Ordered; p = 0.6729 for Random). This demonstrates that whether or not the participants were presented with reputation information in an ordered or random sequence, it did not significantly impact their decision on how much money to send to male faces.

Coefficients:				
	Estimate	**Standard Deviation**	**T-Value**	**Pr(>|t|)**
(Intercept)	**0.7476**	**0.0398**	**18.80**	**<2e-16*** **
Trust Level = 32	-0.1131	0.0613	-1.85	0.0653
Trust Level = 44	-0.0531	0.0517	-1.03	0.3049
Trust Level = 56	-0.0106	0.0528	-0.20	0.8415
Trust Level = 68	-0.0187	0.0459	-0.41	0.6837
Trust Level = 80	0.0225	0.0406	0.56	0.5785
Ordered	0.0254	0.0281	0.90	0.3663
Random	-0.0058	0.0281	-0.21	0.8378

Table 6: Linear regression analysis summary for sent amount based on trustworthiness level and condition in female stimuli.
*Notes: * p <.05, ** p <.01, *** p <.001.*

As we can see from Table 6, when there is no reputation information is provided and the trust level is set to the reference point of 20, which marks the minimum threshold in our analysis, the regression model's intercept suggests a significant initial investment inclination of approximately \$0.75 by participants for female stimuli (p < 2e-16). As the trust level of the female faces increased,

there was no significant effect on the amount invested. This is evidenced by the coefficients and p-values associated with higher trust levels. For instance, akin to the pattern observed at a trust level of 32, where the coefficient is -0.1131 with a corresponding p-value of 0.0653, similar trends in coefficient and p-value magnitudes are evident across other trust levels. This suggests that within the parameters of our study, variations in trustworthiness levels of female faces, beyond the baseline, do not significantly alter the investment behavior of participants.

3.2B. What is the influence of the salience of reputation knowledge on sharing behavior?

Previous regression regressed invested amount of each participant with respect to the baseline condition, however, we need to extend our analysis to include the dynamics introduced by reputation information, specifically focusing on the 'random' versus 'ordered' conditions. Table 6 explores the effect of perceived trustworthiness of the chosen face, reputation score of the chosen face, and presentation type of the faces.

	Estimate	Std. Error	Z-Value	P-Value
(Intercept)	**0.7278**	**8.848e-02**	**8.226**	**1.26e-15*** **
Trust Level	6.260e-04	1.507e-03	0.415	0.678
Ordered	-2.257e-02	1.417e-01	-0.159	0.874
Reputation Score	-4.073e-03	3.059e-03	-1.331	0.184
Trustworthiness * Ordered	2.706e-04	2.214e-03	0.122	0.903
Trustworthiness * Reputation Score	4.857e-05	4.387e-05	1.107	0.269
Ordered * Reputation Score	-4.570e-03	7.526e-03	-0.607	0.544
Trustworthiness * Ordered * Reputation Score	6.119e-05	9.696e-05	0.631	0.528

Table 6: Linear regression summary for sent amount based on the interaction of trustworthiness level, presentation type (excluding baseline and with 'random' as the reference level), and reputation score in male stimuli.
*Notes: * p <.05, ** p <.01, *** p <.001.*

Excluding the predictor variables, participants invested an amount of $0.73 dollars. When we consider the predictor variables individually, none show a significant impact on the amount invested by participants. The 'Trust Level' variable, with a Z-value of 0.415 and a P-value of 0.678, suggests that changes in trust level do not significantly influence the investment decisions. Similarly, the 'Ordered' variable, with a Z-value of -0.159 and a P-value of 0.874, and the 'Reputation Score' variable, with a Z-value of -1.331 and a P-value of 0.184, also fail to show a significant effect on the amount invested.

When we examine the interaction of each predictor on the investment amount, none demonstrate a significant effect. The interaction between 'Trustworthiness' and 'Ordered' has a Z-value of 0.122 and a P-value of 0.903, indicating that the combined effect of trustworthiness and the order of presentation does not significantly affect investment decisions. The interaction between 'Trustworthiness' and 'Reputation Score' shows a slightly more promising Z-value of 1.107 and a P-value of 0.269, but this is still not enough to conclude a significant effect. The same goes for the other interaction terms, such as 'Ordered * Reputation Score' and 'Trustworthiness * Ordered * Reputation Score,' which have P-values of 0.544 and 0.528, respectively, far from indicating any significant influence.

In summary, the linear regression analysis presented in Table 6 suggests that neither the individual variables of trust level, order of presentation, and reputation score, nor their interactions, significantly affect the amount invested by participants in male stimuli.

Table 7 explores the effect of same predictors on investment amount for the female stimuli.

	Estimate	Std. Error	Z-Value	P-Value
(Intercept)	**0.7383**	**0.1111**	**6.647**	**6.9e-11*** **
Trust Level	0.0002	0.0016	0.144	0.886
Ordered	-0.1289	0.1542	-0.836	0.404

Reputation Score	-0.0094	0.0083	-1.131	0.258
Trustworthiness * Ordered	0.0017	0.0023	0.757	0.450
Trustworthiness * Reputation Score	0.0001	0.0001	1.149	0.251
Ordered * Reputation Score	0.0145	0.0102	1.424	0.155
Trustworthiness * Ordered * Reputation Score	-0.0002	0.0001	-1.381	0.168

Table 6: Linear regression summary for sent amount based on the interaction of trustworthiness level, presentation type (excluding baseline and with 'random' as the reference level), and reputation score in female stimuli.
*Notes: * p <.05, ** p <.01, *** p <.001.*

Excluding the predictor variables, participants invested an amount of $0.73 dollars, similar to participants in the male conditions. When we consider the predictor variables individually, none show a significant impact on the amount invested by participants. The 'Trust Level' variable, with a Z-value of 0.144 and a P-value of 0.886, suggests that changes in trust level do not significantly influence the investment decisions. Similarly, the 'Ordered' variable, with a Z-value of -0.836 and a P-value of 0.404, and the 'Reputation Score' variable, with a Z-value of -1.131 and a P-value of 0.258, also fail to show a significant effect on the amount invested.

Interaction of these predictors didn’t result in a significant effect on the investment amount either. The interaction between 'Trustworthiness' and 'Ordered' has a Z-value of 0.757 and a P-value of 0.450, indicating that the combined effect of trustworthiness and the order of presentation does not significantly affect investment decisions. The interaction between 'Trustworthiness' and 'Reputation Score' shows a slightly more promising Z-value of 1.149 and a P-value of 0.450, but this is still not enough to conclude a significant effect. The same goes for the other interaction terms, such as 'Ordered * Reputation Score' and 'Trustworthiness * Ordered * Reputation Score,' which have P-values of 0.155 and 0.168, respectively, indicating a non-significant influence.

4) Exploratory Results

As explained in the hypothesis, the knowledge of reputation could increase the inequality in partner selection due to its presence and how it is presented. To assess the inequality, we calculated linear contrast values for each trial within the conditions. This involved taking the proportions of the accumulated reputation scores for each face (Face 1 to Face 6) and applying a weighting scale from -3 to +3. Specifically, the formula for the linear contrast is as follows:

$$\boldsymbol{linear\ contrast} = -3 \times \boldsymbol{prop}(\boldsymbol{F_1}) - 2 \times \boldsymbol{prop}(\boldsymbol{F_2}) - 1 \times \boldsymbol{prop}(\boldsymbol{F_3}) + 1 \times \boldsymbol{prop}(\boldsymbol{F_4}) + 2 \times \boldsymbol{prop}(\boldsymbol{F_5}) + 3 \times \boldsymbol{prop}(\boldsymbol{F_6})$$

where "prop (F_x)" represents the proportion of the accumulated reputation score for face x. This approach allowed us to quantify the inequality in partner selection based on the distribution of reputation scores across different faces within each trial.

Calculating each linear contrast value for every trial resulted in the following graph:

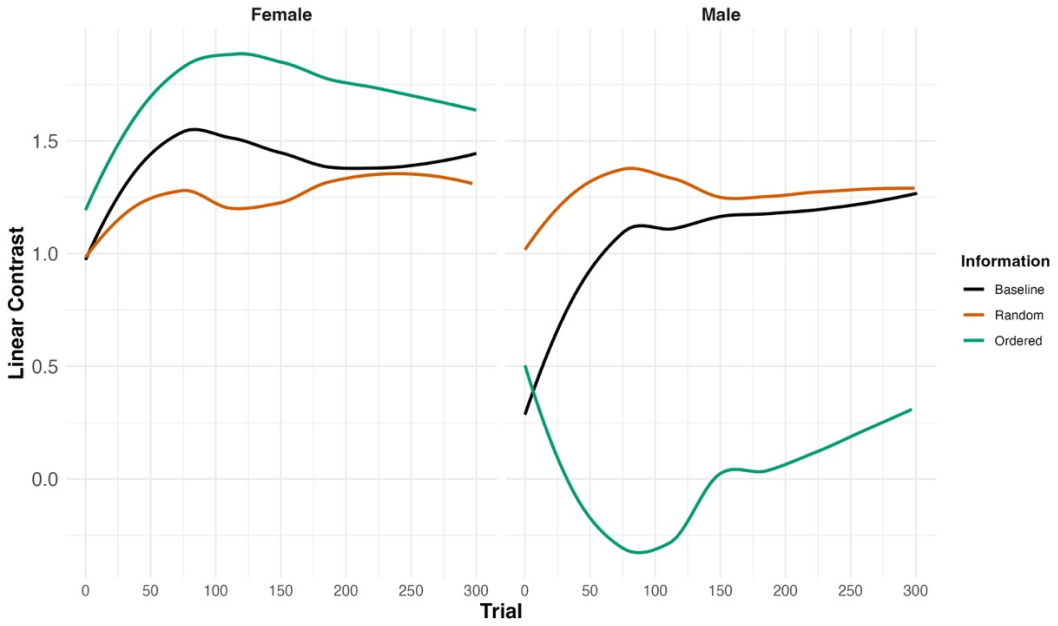

Figure 10: Linear contrast over the duration of trials.

For the conditions involving female stimuli, when participants were not provided with any information about reputation scores, faces deemed more trustworthy initially garnered higher proportions of selections up to about the 80th trial. After that point, other faces began to be chosen more frequently, reducing the linear contrast value until the 200th trial, after which no significant changes were observed. In scenarios where reputation information was randomly presented, serving as a less pronounced social cue, participants tended to choose faces they perceived as more trustworthy throughout the experiment. However, between trials 75 and 125, there was a slight reduction in the likelihood of selecting the most trustworthy faces. When the reputation information was more prominently displayed, with faces ordered by descending reputation scores, participants showed a stronger preference for choosing faces that appeared more trustworthy, more so than when reputation information was less pronounced or absent. Nevertheless, a shift towards selecting less trustworthy-looking faces was noted after the 150th trial.

In conditions with male stimuli, when no information about reputation was available, participants were inclined to select faces they perceived as more trustworthy. This preference was strongest up until the 75th trial, after which it gradually stabilized. With randomly presented reputation knowledge, participants initially favored faces that appeared more trustworthy up to the 100th trial. Following this, there was a noticeable shift away from the most trustworthy-looking faces, with preferences stabilizing after the 150th trial. In the ordered condition, where faces were arranged by their reputation scores, there was an initial trend towards choosing less trustworthy looking faces up until approximately the 75th trial. Subsequently, there was a sharp increase in the selection of faces perceived as more trustworthy.

In summary, the patterns of partner selection varied between different stimuli. Interestingly, while a higher salience of reputation information led to an increased preference for

choosing faces perceived as more trustworthy among female stimuli, the opposite effect was observed for male stimuli. For males, higher salience resulted in a diminished preference for highly trustworthy looking faces.

5) General Discussion

The decision to trust is a complex behavior, reflecting varying patterns in response to the changing dynamics of the signals people are exposed to. Our results supported previous research by showing that people have a tendency to choose higher trustworthy looking faces. This tendency produced significant variation as a function of the information that was provided to the participants for male faces, but it did not for female faces. Thus, we found evidence that when people are presented with direct knowledge that is relevant to the context of decision-making, cooperative behavior is affected. In other words, impressions of trustworthiness are socially influenced by other's choices.

Interestingly, the significant interaction of only one level of trust level in the context of random presentation of reputation knowledge indicates that randomness might have introduced unpredictability, which in turn could make certain levels of perceived trustworthiness stand out more to the observer. Moreover, the lack of significant effects related to the ordered or random presentation of reputation scores overall suggests that the salience of social signals might be less important than the content of the information itself, at least in the context of partner selection based on facial trustworthiness.

Regarding the gender differences, this experiment showed that female faces were perceived differently than male faces, suggesting that this might be due to a gender effect. Nevertheless, partner selection for female stimuli was associated with higher pre-determined levels of trust, and the reputation score of the chosen partner significantly impacted the trustworthiness judgments.

In terms of the willingness to share the initial gift that was given to the participants, there was a higher amount of investment as the trustworthiness of male faces increased, but not form female faces. However, we should note that the observed base high level of invested amount for both stimulus genders indicate an inherent inclination to share, regardless of the trustworthiness perception. Moreover, the salience of reputation knowledge did not significantly affect the propensity to share, even when the reputation score of the chosen face was considered.

Our findings suggest that people's initial biases of perceived trustworthiness can be disrupted by a social signal informing the individual about the specific context or history of interaction associated with the person represented by the facial stimulus. In other words, people tend to reassess their trustworthiness judgments when choosing a partner in a cooperative context.

References

Albohn, D. N., Martinez, J. E., & Todorov, A. (2023, November 28). Determinants of Shared and Idiosyncratic Contributions to Judgments of Faces. https://doi.org/10.31234/osf.io/78jez

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences*, *104*(46), 17948 17953.

Bromberg-Martin, E. S., & Sharot, T. (2020). The value of beliefs. *Neuron*, *106*(4), 561-565 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.

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.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, *55*, 591-621.

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.

Ewing, L., Caulfield, F., Read, A., & Rhodes, G. (2015). Perceived trustworthiness of faces drives trust behaviour in children. *Developmental science*, *18*(2), 327-334.

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.

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior, 34*(2), 83 110. https://doi.org/10.1007/s10919-009-0082-1

Olivola, C. Y., Eubanks, D. L., & Lovelace, J. B. (2014). The many (distinctive) faces of leadership: Inferring leadership domain from facial appearance. *The Leadership Quarterly*, *25*(5), 817-834.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087-11092.

Peterson, J. C., Uddenberg, S., Griffiths, T. L., Todorov, A., & Suchow, J. W. (2022). Deep models of superficial face judgments. *Proceedings of the National Academy of Sciences*, 119(17), e2115228119.

Porter, S., Ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law*, *16*(6), 477-491.

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in cognitive sciences*, *17*(8), 413-425.

Salganik, M. J., & Watts, D. J. (2008). Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market. *Social psychology quarterly*, *71*(4), 338-355.

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *science*, *311*(5762), 854-856.

Semmann, D., Krambeck, H. J., & Milinski, M. (2005). Reputation is valuable within and outside one's own social group. *Behavioral Ecology and Sociobiology*, *57*, 611-616.

Sharot, T., Rollwage, M., Sunstein, C. R., & Fleming, S. M. (2023). Why and when beliefs change. *Perspectives on Psychological Science*, *18*(1), 142-151.

Todorov, A., & Oh, D. (2021). The structure and perceptual basis of social judgments from faces. In *Advances in experimental social psychology* (Vol. 63, pp. 189-245). Academic Press.

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. https://doi.org/10.1016/j.cognition.2008.07.002

Van de Rijt, A., Kang, S. M., Restivo, M., & Patil, A. (2014). Field experiments of success breeds-success dynamics. *Proceedings of the National Academy of Sciences*, *111*(19), 6934-6939.