Face-off: Facial Features and Strategic Choice[1]

Dustin Tingley[2]

This draft: April 5, 2012

[1]My thanks to Irving Dominguez, Michael Gill, Eva Ghirmai, and Peter Knudson for research assistance, the Todorov team for making available the avatar images, Lisa Camner, Ryan Enos, Alice Hsiaw, Joo-A Julia Lee, Rose McDermott, Kristin Michelitch, Chris Olivola, Chris Said, Michael Tierney and participants in the Harvard Government Department Political Economy seminar for discussions, and the Harvard Decision Science Laboratory for use of their facilities. All mistakes are my own.

[2]Government Department, Harvard University, Cambridge, MA 02138. Email: dtingley@gov.harvard.edu,

Face-off: Facial Features and Strategic Choice

Abstract

I study experimentally a single shot trust game where players have the opportunity to choose an avatar–a computer generated face–to represent them. These avatars vary on several dimensions–trustworthiness, dominance, and threat–identified by previous work as influencing perceptions of those who view the faces (Todorov *et al.*, 2008). I take this previous work and ask whether subjects choose faces that are ex ante more trustworthy, whether selected avatars have an influence on strategy choices, and whether individuals who evaluate faces as more trustworthy are also more likely to trust others. Results indicate affirmative answers to all three questions. Additional experimental sessions used randomly assigned avatars. This design allows me to compare behavior when everyone knows avatars are self-selected versus when everyone knew they are randomly assigned. Random assignment eliminated all three effects observed when subjects chose their avatars.

1 Introduction

Trust has long been recognized as foundational to political and economic interaction (Levi and Stoker, 2000). Trust underlies the success of democratic institutions (Mishler and Rose, 1997), influences Presidential evaluation (Hetherington, 1998), and is even taken as indicator of political disengagement (Putnam, 2000). While a range of important work highlights the ways social and political institutions influence trust (e.g., Anderson, 2010), this paper focuses on the role of physical appearance in determining trustworthiness. A range of research in psychology, economics, and political science suggests that physical appearance plays a very strong role in influencing our behavior. Not only are earnings within firms predicted by perceptions of both physical attractiveness and competence (e.g., Biddle and Hamermesh, 1998), but election outcomes are also predicted by these variables (Ballew and Todorov, 2007; Atkinson *et al.*, 2009; Mattes *et al.*, 2010; Olivola and Todorov, 2010a). Even rank attainment in the US military has been related to physical features of officers (Mazur *et al.*, 1984). More broadly, the ability to detect intentions to cooperate plays a key evolutionary function, and given the strong role played by the human face in this process and the neural systems that play a strong role in evaluating faces, it is likely that faces play a key role in decoding intentions (Chiappe *et al.*, 2004; Oda, 1997; Wentura *et al.*, 2000; Winston *et al.*, 2002; Yamagishi *et al.*, 2003).

In this paper I explore the influence of facial features on trusting behavior in an financially consequential context. I introduce a variation to existing experimental designs that has subjects choose computer generated avatars to represent them in the experiment. These avatars were specially manipulated to vary along dimension that previous research shows has an important influence on on perceptions of intentions. Having subjects choose faces to represent them in a strategic setting is a novel design and interesting for several reasons. First, it allows us to check if subjects have an intuitive notion of what intentions faces are likely to communicate as well as provide an financially consequential test of previous work that identifies variation in how faces are perceived. Second, while a range of studies document the influence of individual appearance on elections and hiring decisions, political parties and employers–who might be thought of as principals–have control over who they choose to represent their interests. These decisions may be based on appearances in order to have agents that are maximally effective. Letting individuals choose avatars to represent them parallels this situation, albeit in a highly stylized and controlled laboratory setting. For example, if one combines the findings by Abramowitz (1989) that primary voting is influenced by (or correlated with) electability, and the range of findings that appearance influences electability (Ballew and Todorov, 2007; Atkinson *et al.*, 2009; Mattes *et al.*, 2010; Olivola and Todorov, 2010a; Lawson *et al.*, 2010), ceteris paribus, core party supporters may have an incentive to choose candidates to represent them based upon their physical appearance. Similar logics extend to the workplace, where employers might prefer individuals to look a certain way to manipulate the external appearance of the firm to outside clients. In a follow-up experiment reported below, I had participants choose from the same set of avatars used in the main experiment who to send as a mediator between two disputing parties, where it was important to have the disputing parties trust the mediator. Third, studies of cheap-talk frequently use verbal or written communication as the medium for communication. However, selection of avatars permits communication using a non-verbal "grammar." Earlier results discussed below suggest that humans might have a common comprehension of how different faces project different intentions and the present study examines this proposition directly.

The design of the experiment falls out of two recent research programs in psychology and economics. First, recent work suggests evaluations of human facial features vary along several dimensions including trustworthiness (a valence dimension associated with approach/avoidance intentions), dominance (a hierarchial/power dimension associated with strength), and threat (a combination of the first two dimensions) (Oosterhof and Todorov, 2008; Todorov *et al.*, 2008). This results has emerged from studies in which subjects are asked to evaluate pictures of faces and score them along various dimensions. What has not been studied systematically using these dimensions is how they relate to actual economic choice. The present study uses "avatars," computer generated faces that vary along these dimensions, from Oosterhof and Todorov (2008) to represent decision-makers in the experiments. In the experiments presented here subjects hear a description of the strategic game (a single-shot trust game),[1] choose avatars to represent them in the game conditional on their position in the game, and then play the game observing both their opponent's avatar selection and their own. In additional experiments the avatars were assigned randomly.

A second literature, this one from economics, investigates how subjects in interactive experiments behave while playing the well known "trust game" when they can observe a picture of a partner's face. The importance of trust for understanding social and economic interactions has been emphasized across a range of literatures (e.g., Kydd, 2007; Eckel and Petrie, 2011). Scharlemann *et al.* (2001) argue that "(f)or an individual, the key to successful cooperation is the ability to identify cooperative partners. The ability to signal and detect the intention to cooperate would be a very valuable skill for humans to posses" (pg. 617). To study this signalling behavior, they examine the role of smiling in a two-person modified one shot trust game. In the experiment subjects first had their pictures taken, one with a a neutral expression and one smiling. These pictures were then used to represent the subjects in the experiment. They found that when individuals were shown a smiling face of their partner subjects were more likely to trust their counterpart. Separate evaluation of the faces showed that faces that loaded heavily onto a dimension the authors labeled "cooperative" (using a semantic differential survey) were also more likely to be trusted than faces that loaded weakly on this dimension.[2] More recently, Eckel and Petrie (2011) examine behavior in a trust game where subjects in some treatments had the opportunity to pay to see a photograph of their opponent. They show

[1]In the standard trust game formulation (Berg *et al.*, 1995) a sender is given an allocation of money. They can send some amount of it to a receiver, and this amount is increased by some positive scalar. The receiver then decides how much to return.

[2]Scharlemann *et al.* (2001) note that the cooperative dimension is correlated with smiling but more strongly predicts behavior than a smile alone.Van't Wout and Sanfey (2008) report a similar study where individuals play a trust game with a photograph of their partner's face. Individuals who scored as looking more trustworthy in a pre-experiment were sent more money than those with lower trustworthy scores. Other research explores the role of facial similarity. For example, DeBruine (2002) use a sequential trust game to investigate how the amount someone sends to the trustee depends on how much the trustee resembles the sender's own facial features.

that people assign economic value to information about opponent faces, and furthermore that this increases social efficiency. Research like this and others (Frank, 1988; Engell *et al.*, 2007; North *et al.*, 2010; Todorov *et al.*, 2010; Stirrat and Perrett, 2010; Rezlescu *et al.*, 2012) provides evidence that human facial features can signal social intentions (e.g., trustworthiness).[3] Trust plays a crucial role in a number of political contexts, discussed above, and so furthering our understanding of what influences perceptions of trustworthiness is important not just for political science but also for cognate fields.

Building on this previous work, I explore new questions about the relationship between facial features and strategic interaction. First, if humans have an intuitive sense of what types of faces signal particular intentions, then what faces would subjects choose to represent themselves, given the economic context they face? In particular, if subjects were to play a trust game, would the majority of subjects choose faces that are more trustworthy in appearance? Second, what are the underlying dimensions on which faces are evaluated, and do these dimensions play a similar or dissimilar role across different contexts? In particular, are evaluations of trustworthiness most salient in the trust game, or are other dimensions of facial characteristics like dominance and threat more relevant? Third, do people who tend to evaluate faces as being more trustworthy also tend to treat those persons in a more trustworthy manner? This question has two components. First, is variation in trusting behavior across individuals in part attributable to variation in how an individual perceives faces and hence decodes the likely intentions of others? And will individuals be influenced by ostensibly "cheap talk" signals of trustworthiness based on non-verbal signals of avatar choice. Whether cheap talk has any influence on behavior is an important question in politics (Austen-Smith, 1990) and is well suited to experimental investigation (Crawford, 1998; Tingley and Walter, 2011).

The new experimental design I deploy has subjects choose avatars to represent them in the experiment. The design generates some interesting findings. In the trust game, subjects regularly chose avatars that were more trustworthy in appearance, even though they were given no information about the faces. This provides new evidence for the

[3]Other important research documents the various ways that social context can influence trust (Berg *et al.*, 1995; Haley and Fessler, 2005; Delgado and Phelps, 2005).

intuitive understanding of humans about what constitutes a trustworthy face (Todorov *et al.*, 2008) in a monetarily consequential environment (see also Rezlescu *et al.* (2012)). I also find that the sender's perceptions of receiver avatar trustworthiness positively influences the amount of money sent. Individuals who perceived a selected avatar as more trustworthy sent the receiver more money. This suggests, preliminarily, that individuals who evaluate faces to be more trustworthy are also those who exhibit more trusting behavior. It is possible, though not definitively shown here, that individual variation in trust behavior is due to differences in how faces are perceived. If true, this suggests that conventional accounts stressing cultural or experiential explanations for individual variability in trust are incomplete. Finally, evaluations of trustworthiness exert greater influence on behavior compared to evaluations of dominance or threat, an intuitive but heretofore undocumented relationship. The paper proceeds as follows. Section 2 lays out the theoretical ideas in more detail, Section 3 describes the experimental design, Section 4 presents the empirical results, and Section 5 concludes.

2 Physical Appearance and Trustworthiness

Humans rely heavily on physical cues to guide them in how they interact with other individuals. This means that politics, which is inherently social, may be importantly influenced by physical cues. One rationale for the reliance is that physical cues, and in particular expressions or structural features of the face, are informative of another's intentions or dispositions (Yamagishi *et al.*, 2003; Oda, 1997). For example, Frank (1988) argues that expressions rely on relatively automatic neural processes and so people have a hard time "lying" about their intentions. Scharlemann *et al.* (2001) follow this up and show how smiling pictures of participants in a trust game lead to more trusting behavior. Other research (Oosterhof and Todorov, 2008; Todorov *et al.*, 2008) documents how people perceive structural features of the human face along approach/avoidance and dominance dimensions. The authors argue that the approach/avoidance dimension signals trustworthiness. Finally, they investigate a combination of the two dimensions which they label as threat. Todorov *et al.* (2010) and others (Willis and Todorov, 2006) find that these impressions are made even following exposure to a face for very short time period. Furthermore, there is substantial evidence that specific regions of the brain are involved in processing faces, indicating a specialized functional adaptation (e.g., Kanwisher, 2010; Winston *et al.*, 2002). Just as Frank (1988) theorizes that expressions might be able to signal intentions, it makes sense that more fixed features of the face (Said *et al.*, 2009) as well as individual evaluations of face trustworthiness may influence choice behavior.[4] The implication for the study of politics is that physical features, in particular the human face, could influence the selection of candidates or other political agents. Indeed, recent work in political science shows such a connection (Lawson *et al.*, 2010).

While individuals can take on a range of expressions, often depending on their temperament at the time, Scharlemann *et al.* (2001) document the role for more fixed features of the face. In their study, neutral expressions on faces were evaluated along a range of descriptives with subjects choosing from sets of paired words that they thought best described the face. A factor analysis pulled out underlying components of these evaluations. One such component was labeled as "cooperative" which included loadings on friendly/unfriendly, cooperative/non-cooperative, forgiving/unforgiving, happy/sad, and amiable/hostile. In strictly non-economic settings, Todorov *et al.* (2008) and the research they review identify dimensions of the face using a data driven approach. Subjects evaluated faces across a range of words and a principal component analysis extracted the underlying dimensions of the evaluations. Evaluations of trust and dominance loaded most strongly on the two most salient dimensions. They argue that the ability to evaluate faces along these dimensions drives inferences about behavioral intentions (such as trustworthiness).[5] These findings suggest that if individuals in a trust game were able to choose faces to represent them, we should expect them to choose more trustworthy looking faces–those that signal approach rather than avoidance intentions. This is because their partner also holds an intuitive conceptualization of what intentions these faces confer.

[4]Some recent research suggests that people in fact rely on these physical cues more than they should and at the expense of other information (Olivola and Todorov, 2010b).

[5]The authors also explore how perceptions of dominance link with the establishment of power hierarchies.

On this account, individuals create expectations about the behavior of their opponent's future behavior (Ashraf *et al.*, 2006) based upon a choice of avatar. This is an important observation because it helps us understand how people form expectations. In situations of repeated observation and exchange reputational dynamics are likely to trump cheap talk statements (Bracht and Feltovich, 2009). But "first impressions" are an important part of social interaction as well. A large literature–from which the work of Todorov's team stems–considers the determinants of these initial impressions (Willis and Todorov, 2006; Olivola and Todorov, 2010b). In the context considered here, it is possible to try to explore what shapes these expectations. Given the possibility that structural features of the face (as opposed to simply expressions) can signal intentions (Said *et al.*, 2009), we might expect senders to infer intentions from choice of avatar. Signalling intentions is important in politics because citizens/principals cannot always monitor the behavior of political agents. Efforts to understand the intentions of others in the trust context are particularly important in light of evidence about the role of betrayal aversion (Bohnet *et al.*, 2008). A receiver who chooses a more trustworthy looking face might be trying to signal that he/she can be trusted. If senders believe that people will choose faces that match their intentions, then they could condition the amount they send based on this signal. Alternatively, senders might recognize that these signals are not credible and hence dismiss them, just as cheap talk might be dismissed. While most work on cheap talk considers verbal or written forms of communication, modulation of physical cues could serve a signalling role as well. In the experiment below I estimate whether subjects in fact infer anything about intentions based on avatar choice.

A final theoretical starting point is the question of why some individuals are more trusting (as opposed to trustworthy) than others. Earlier work by Glaeser *et al.* (1995) shows that individuals with generalized tendencies to trust others in society are also more trusting in laboratory trust games. While this helped establish greater external validity for laboratory experiments and several methodological points, broader questions are also at play. Following others in the political science literature (e.g., Putnam, 1993), they stress that the density of an individual's social network is highly predictive of trusting behavior. The question of interest here is whether people who evaluate individual faces as being more trustworthy also choose more trusting strategies. For example, consider two people, Bob and Frank, who share similar cultural and life experiences. If Bob finds face X more trustworthy than Frank, would Bob send more money to a person with face X? Perhaps one basis for variation in trusting is that, in fact, more trusting people also decode the faces of others in ways that make them hold more trust in others. Bob behaviorally trusts the person with face X more because he perceives face X as signalling trustworthy intentions. Put differently, some people tend to trust others more simply because they perceive the faces of others as being more trustworthy. While this of course does not explain why people decode faces in different ways, it suggests that attributions of variation in trusting to broader social forces–such as social capital or density of social networks–might be premature or at least not the entire story. In this sense I provide a preliminary exploration of how differences in facial evaluation explain behavior alongside the "cultural/experiential" type explanations (Pinker, 2002) common in conventional accounts. The present study by no means verifies that this conjecture is true, but I do present some preliminary evidence to this end.

3 Experimental Design

3.1 Avatars

In the primary experiment subjects first learned about the structure of the trust game and then chose a face to represent them from a set of computer generated head shots taken from Todorov's library. A primary reason to use these faces, rather than some other set of faces including real ones, is that the exact way that they vary is more tightly controlled and hence there is less risk that other dimensions of the faces will drive our results. This, of course, produces trade-offs with other concerns.[6] There were two sets of faces used in the experiment. The first set of faces are based on a linear model predicting levels of trust,

[6]For example, in a student sample people might have more familiarity with the use of "avatars" from their experience in various virtual world experiences.

dominance, and threat that earlier research had developed.[7] The faces are generated via the FaceGen 3.1 software using the procedures outlined in Oosterhof and Todorov (2008). Throughout the paper these are referred to as the "*generated*" faces. The faces used in the present experiment were selected from a single set of faces in the Todorov data-set (set 1) and appear in Table 1. The selected faces represent -2 standard deviations from the mean, the mean, and +2 standard deviations for each dimension (trust, dominance, and threat). More extreme faces were not used because faces at these extremes are no longer emotionally neutral (Todorov *et al.*, 2008, pg. 457).[8] While the mean faces are not identical[9] they are more similar than other sets that others generated from the same base regression model. For the trust dimension faces, a higher number indicates a more trustworthy face, and for the dominance and threat dimensions a higher number indicates a more dominating or threatening face. Like a parallel study (Rezlescu *et al.*, 2012), use of these faces have the advantage of avoiding potential confounds present in other studies that use real faces (e.g., Van't Wout and Sanfey, 2008).[10] At no point were subjects told anything about the origin of the faces or the typologies they represent. It is important to point out that all of the faces in fact vary along all three dimensions of trust, dominance, and threat.

A second set of faces that subjects selected from (though not at the same time as the "generated" faces) is displayed in Table 2 and are referred to as "evaluated" faces. These faces are also computer generated, but instead of being based on a regression model they were evaluated by human subjects across a number of dimensions, including how

[7]An extended discussion of the relationship between these dimensions and structural features of the face as well as perceptions of emotional dispositions is elsewhere (Todorov, 2011).

[8]In this sense the present study represents another important departure from Scharlemann *et al.* (2001) who were interested in the role of smiles.

[9]In a personal correspondence with Todorov's team, they indicated that this was not possible.

[10]The present study differs from Rezlescu *et al.* (2012) in important though complementary ways. I investigate choice of faces in a no-deception environment whereas they assign faces but tell subjects they actually represent their partners. They also use manipulations of trust that are more extreme (+/- 3 s.d.) and in additional experiments explore the intersection of providing information (again manipulated by the researchers) about the behavioral history of a partner. They show that even with this history more trustworthy Role 2 faces receive more.

trustworthy they look. Across the 300 pictures in this data set provided by the Todorov team, I selected 9 in order to simplify the choice task for the subjects. In the entire 300 picture data set the lowest trustworthy score was 2.9, the highest 6.4, and the mean was 4.8. The 9 faces I use range from 3.4 to 6.1 with a 4.7 average. I label the faces TR1 to TR9 in descending order of trustworthiness. These faces also varied along the dimensions of dominance and threat, and were also more heterogenous in terms of things like skin tone. These faces were used to help probe the robustness of this new type of research design by utilizing a greater variety of face types.

In the experiments, subjects chose from either the generated or evaluated faces, with some sessions using the generated faces first and the evaluated faces second, and other sessions reversing this order. I used the two different sets of faces for several reasons. First, given that no previous research has used the scaled faces this way (Oosterhof and Todorov, 2008; Todorov *et al.*, 2008), there is little ex ante reason to suspect the generated or evaluated sets is preferable to the other. If future research uses controlled variation of these faces then it would be helpful to understand whether faces with more controlled variation should be used or whether using pre-existing evaluations is better. Second, while the generated set provides clearer ex ante scaling along the dimensions of interest, they were generated via a regression model and hence only can be expected to vary along these dimensions in expectation; actual human evaluation could differ. Conversely, the "evaluated" set already went through a process with human coders evaluating the faces along a number of dimensions including trustworthiness, dominance, and threat. Finally, it is possible that idiosyncracies in one versus the other could bias the results and so I study both.

3.2 Experimental Game

The trust game is a widely studied game with the following structure (Berg *et al.*, 1995). An individual in Role 1–the "sender"–can choose to send some amount of money provided in an initial allocation ($x \in [0, 50]$ in the current experiment) to the person in Role 2–the "receiver." This amount is then increased by some scalar $k > 1$ ($k = 3$ in the current experiment). For example, if the sender chose to send 20 points the receiver would get 60.

Finally, the receiver chooses how much of $k \times x$ to return, denoted z, and then keeps the remaining amount. Payoffs respectively to Role 1 and Role 2 players are $(50-x+z, kx-z)$. The standard Nash equilibrium prediction for the model is for the sender to keep the entire initial allocation and the receiver to keep any amount sent. In practice, as a number of studies have shown, the amounts sent and returned are greater than 0.

3.3 Procedures

Experimental sessions were run in the Harvard Decision Science Laboratory (HDSL) using computer workstations with blinders. Subjects were undergraduate students registered with the HDSL subject pool. Subjects participate in 6 repetitions of the experiment. In each repetition, all Role 1 (sender) subjects are paired with all Role 2 (receiver) subjects once and in a random order. Hence if there are ten subjects, I observe five plays of the trust game in a single repetition of the experiment. Prior to each repetition subjects were randomly assigned either to Role 1 or Role 2 and chose which avatar to represent them. For several experimental sessions, subjects chose from the "generated set" in the first 3 repetitions. In repetitions 4-6, the "evaluated set" was used. In other sessions this order was reversed to control for potential ordering effects. Once the participants were matched, the Role 1 and Role 2 avatars were displayed on the left hand side of the screen, and the pair would play a one-shot trust game. All interactions were anonymous. In the experiment, the choice of the Role 2 person of how much to send back was not displayed to the Role 1 person in order to limit population based learning.[11] Points in the experiment were converted to money at 10 points = \$1. To pay subjects, a randomly determined pairing from a randomly determined repetition (out of 6) was chosen. Hence subjects were paid based on either a Role 1 or Role 2 position. All information was common knowledge and subjects were paid privately at the end of the experiment.

After completing the trust game part of the experiment, all subjects completed a

[11] Such dynamics are not of interest in the current study. The goal of the experiment was to isolate the influence or Role 2 (receiver) avatar choice on Role 1 (sender) choice of how much to send. While isolating this influence makes the results less ecologically valid, this is the correct design choice given the hypotheses under investigation.

survey that measured several demographic variables, personality scores, and evaluations of the faces used in the experiment. In the evaluation section subjects rated generated and evaluated faces on how trustworthy, dominant, and threatening they looked on a $1-7$ scale. The order of faces within each set was randomized. These evaluations permit examining the relationship between the level of trustworthiness Role 1 perceives in Role 2's face and the amount that Role 1 sent to Role 2 in the actual experiment. Importantly, in the trust game section of the experiment the amount returned was never revealed and so individual could not form expectations about particular faces that could bias these evaluations. In addition, individuals in most sessions were asked hypothetically how much they would send to each of the 18 avatars.

The number of subjects per session was 8,10, or 12, for a total of 60 total subjects in six sessions.[12] In four of the sessions subjects chose from the "generated" faces for the first three repetitions and in two (each with 12 subjects) the "evaluated" faces were chosen from in the first three repetitions. I also ran an additional four experimental sessions with 40 different subjects where the avatars were randomly assigned and this was commonly known. In three of these sessions the generated faces were used first and in one the evaluated set was used first. This randomization helps separate the effect of the avatars being present in general from from any inferences made about the trustworthiness of the Receiver based on their choice of avatar.

3.4 Hypotheses

In the experiments, Senders and Receivers were able to choose avatars to represent them in a single-shot trust game. The data contain which avatars were chosen, how individuals and the group as a whole evaluated the characteristics of the faces, and choice behavior (amount sent and returned). I explore several hypotheses motivated in the preceding sections.

[12] The number of subjects per session varied because sometimes subjects failed to show, and sessions require an even number. Subjects were not told the total number of subjects participating in the experiment.

Hypothesis 1: Subjects will be more likely to choose avatars to represent themselves that score higher in trustworthiness and/or lower in levels of threat and dominance.

Hypothesis 2: The amount sent to a receiver who chose an avatar with a higher *average* trustworthy rating will be greater than the amount sent to a receiver who chose an avatar with a lower *average* trustworthy rating.

Hypothesis 3: Senders will send larger amounts when they *individually* perceive the receiver's avatar as particularly trustworthy, according to the sender's *own* post-experiment evaluation, and less when the chosen avatar is perceived as less trustworthy.

Hypothesis 4: Any influence of Role 2 avatars on Role 1 choices will be eliminated if avatars are randomly assigned.

Hypotheses 2 and 3 are clearly related. However, they differ in the sense that Hypothesis 2 is about difference in behavior that depends on *average* differences in evaluations of an avatar (using the post-experiment evaluations) whereas Hypothesis 3 is about individual level differences in evaluations of avatars. In particular, individuals might have slightly different perceptions of how trustworthy a particular face is. Hypothesis 3 picks up on this possibility and allows for greater individual level variability in perceptions of trustworthiness, whereas hypothesis 2 only tests the influence of each avatar's average trustworthiness score. Hypothesis 4 suggests that the mechanism that produces the effect described in Hypotheses 2 and 3 operates via the transmission of information about intentions. In principle such information is "cheap talk," albeit communicated through facial features as opposed to language. Removing this possibility for communicating by randomly assigning avatars should eliminate any cheap talk effects, if they are present.[13]

[13] An additional hypothesis suggested by several readers is that the amount sent should be higher if both chose the same face. I tested this hypothesis and did not find support for it.

4 Analysis

4.1 Choice of Avatar

I begin with Hypothesis 1 and the choice of avatars by those in the Role 2 (receiver) position. The top row of Figure 1 presents sessions where avatars were chosen and the bottom row for sessions where the avatar was randomly assigned. In the latter category random assignment is evident given the approximate uniformity of the distribution. Above each avatar option is the average trust score from a post-experiment survey. In sessions where avatars were chosen from the generated set, subjects predominantly chose the TW5 face (+2 sd on trustworthy dimension) and the Threat1 face (-2 sd on threat dimension). The high frequency of TW5 choices provides the strongest support for hypothesis 1. Amongst the 222 cases where TW3 or TW5 were chosen, 67% of cases were TW5, which is a significantly different proportion compared to .5 ($p < .01$). Furthermore, while the proportion of the highest trust dimension avatar and lowest threat dimension avatar (Threat1) were statistically indistinguishable, compared to other dimensions they each had significantly higher proportions. The high frequency of choices of the Threat1 face is also consistent with Hypothesis 1 because the threat dimension is a combination of the trust and dominance dimensions (Todorov *et al.*, 2008). Hence we should expect low threat to also have high perceived levels of trust. Indeed, as shown in Figure 1 the average trust scores of the two avatars in the post-experiment survey are nearly identical. Thus while there is strong support for Hypothesis 1, this is qualified in that low threat is extremely correlated with high trust. The other avatars were less frequently chosen.[14]

Choices from the evaluated avatar set reflect a similar pattern, with Role 2 choices most frequently being faces rates most trustworthy in previous experiments. Individuals in the Role 2 position choose faces that attempt to signal a trustworthy presence. Hy-

[14]The infrequent selection of the low dominance face is interesting, illustrating the orthogonality of the dominance and trust dimensions. Were low dominance simply akin to trust, then we would expect this face to be chosen with similar frequency as the high trust or low threat (a combination of low dominance/high trust) faces. This is not the case.

pothesis 1 receives strong support.[15] Furthermore, the results with the generated faces provide additional nuance to and support of the trust dimension identified by the Todorov team, but here in reference to choices that were part of an interactive economic game.[16] Interestingly, choice of avatars in the Role 1 position (Figure 2) look strikingly similar to the Role 2 position choices. Apparently in the trust game individuals in both the Role 1 and Role 2 positions gravitate towards similar avatars to represent them.

These results suggest that individuals have an intuitive understanding of how physical features of the human face would be interpreted in a particular incentive context. In politics, individuals may thus evaluate what type of incentive problems they face and choose political candidates, or agents, based in part on physical appearances. Indeed, in Section 4.4 we explore this connection directly.

4.2 Amount Sent

Next, I investigate the amount sent by the Role 1 person with respect to Role 2's chosen avatar. The top row of Figure 3 plots the mean and 90% confidence intervals of amounts sent for the generated and evaluated face sets. For the generated faces several avatars were never chosen (Dom3, Dom5, and Threat5). Other avatars, such as TW1, were chosen relatively infrequently and have large confidence intervals around the mean. Importantly, the mean amount sent to a Role 2 player with the high trust avatar, TW5, is higher on average than the amount sent to those selecting the TW3 avatar. This mean difference is statistically significant ($N = 222, abs(t) = 1.95$) as were rank based tests. Role 2 subjects selecting the TW5 avatar, which is scaled to be 2 standard deviations greater on the trust dimension than TW3, received a higher average amount from their Role 1 partners. This is perhaps especially striking given the similarity of the faces. This indicates a greater

[15] Some subjects did not choose the most trustworthy looking faces. Future work might consider the individual determinants of these choices.

[16] In additional experiments with separate subjects an Ultimatum game was used. Figure 6 of the online appendix displays avatar choice frequencies for these experiments. Role 2 choices in the ultimatum game (the accept/reject decision-maker) look substantively different, with a higher frequency of mid to high dominant and threat faces, as well as more low trust faces chosen. Additional details discussed in Appendix 6.1.

degree of trust in the Role 2 person, the only difference being which face was chosen. For the evaluated faces, the commonly chosen TR1 and TR3 avatars received higher amounts sent than several other avatar types. This is especially clear when looking at the medians. These differences were not regularly significant, though.

While these plots give some sense of the distribution of amounts sent, they do not definitively show that the amount sent is correlated with the trustworthiness of the Role 2 avatar. To explore the relationship between the amount sent and the trustworthiness of the avatar we must take into account several additional factors. First, some of the data is censored at 0 and at 50. Role 1 subjects could transfer all or none of the resource. I use tobit regression to take this into account. Second, the avatars from the experiment were selected to vary on the dimensions of trust, dominance, and threat. However, individual subjects may evaluate the faces in different ways. For example, in the generated set, faces were created using a regression model. This suggests that while on average faces with greater trustworthiness should be viewed as more trustworthy, individual participants may evaluate faces differently. This issue is potentially less severe for the evaluated faces, in that those faces were actually scaled by subjects in Oosterhof and Todorov (2008) earlier studies. However, there still could exist individual differences that should be taken into account. In the post-experiment survey, I measure evaluations of each face along the trustworthy, dominance, and threat dimensions. In addition, the order in which the avatar sets were used–generated first versus evaluated first–may impact behavior in the game. Finally, subject specific characteristics could influence choices.[17] In what follows, I consider all of these nuances by moving to multivariate models.

I explore several different ways to test the influence of how trustworthy a Role 1 person finds Role 2's selected face. First, for each avatar, I calculate the average trustworthiness evaluation across subjects measured in the post-experiment survey. I then merge this score into the data for both the avatar chosen by the Role 1 person (*AvgTrustOwn*) and the Role 2 person (*AvgTrustOther*). Thus for each play of the game I know how

[17]For example there is evidence that there exists a correlation between general attitudes toward trusting others in society and behavior in a trust game (Glaeser *et al.*, 1995). While random pairing of subjects would mitigate a bias, the above results ignore these difference.

trustworthy on average subjects found the Role 1 and Role 2 avatars. This permits a test of hypothesis 2. Second, I use each individual's own post-experiment trustworthiness evaluation of each avatar and merge these values into the data according to which avatars were actually chosen by the player and his/her opponent, producing *IndvTrustOwn* for Role 1 and *IndvTrustOther* for Role 2's avatar. Higher values of all of these variables indicate a higher trustworthiness rating. This permits a test of hypothesis 3.

I use a tobit regression model and cluster robust standard errors at the subject level to account for correlated choices within subjects. Table 3 presents the results. The first three models include all observations where subjects chose avatars. I control for repetitions using generated faces (1) versus evaluated faces (0), (*Generated*), for the repetition of the experiment (*Repetition*, six per session), a response to a general level of trust question (*WVSTrust*)[18], and each individual's average trustworthiness ranking across all avatars (*AvgTrust*).[19] The coefficient on *AvgTrustOther* is positive but not significant, which is consistent with Figure 3 because we are pooling across Generated and Evaluated faces. However, moving to individual level trustworthiness scores, *IndvTrustOther*, which Hypothesis 3 suggests should be influential, we find more supportive evidence. The coefficient is positive and significant whether or not the sender's perceived trustworthiness of their own avatar is controlled for. In model All2, a one unit change in how trustworthy an individual finds the receiver leads to an additional 3.2 points sent.

The second set of models consider only generated faces. Here we see stronger support for hypotheses 2 and 3. The coefficient on *AvgTrustOther* has one-sided p-value of .06, with a one-sided test being reasonable given that the stated hypotheses are clearly one directional. Looking next to the models with individual level evaluations of avatar trustworthiness, we observe a positive and significant coefficient for the evaluations of the

[18]How much do you agree with the following statement: "Most people can be trusted." Scored along a 1 to 5 scale with 1 disagree completely and 5 agree completely. Question taken from World Values Survey.

[19]In additional robustness checks I also included an indicator variable for whether Role 1 and Role 2 chose the same face. This variable was positive but never significant and did not change the results reported here. A control variable for gender of the participant was also never significant and did not influence the results.

Role 2 avatar (*IndvTrustOther*) whereas the *IndvTrustOwn* coefficient is not significantly different from 0. The magnitude of the effect of the *IndvTrustOther* is an increase in amount sent of 3.6 for a unit change in the explanatory variable. The third set of models uses only evaluated faces. Here we see the weakest results, though *IndvTrustOther* is still significant in these models.

These results hold even when including control variables for how trustworthy someone believes "most people" are (*WVSTrust*) and each subject's average trustworthiness evaluation across all avatars (*AvgTrust*). Even when controlling for general dispositions to trust others, a tendency to evaluate other's faces as more trustworthy on average, and the perceived trustworthiness of one's own avatar, individuals send more to individuals with avatars with greater perceived trustworthiness. This provides preliminary evidence that senders used the faces as a signal of the receiver's intent. This shows that subjects share a common, intuitive, understanding of the information contained in the selected faces. It also suggests that "cheap-talk" does have an influence on choice.

These results suggest that the choice of how much to send to Role 2 is a function of how an individual perceives a face and processes this information to form a view of the other person's intentions. This further suggests that while cultural and experiential variables (Glaeser *et al.*, 1995) are likely to be important, they might not be the whole story in explaining variation in trust across individuals. Here, all subjects were drawn from a similar subject pool and the *WVSTrust* is a reasonable summary statistic for the effect of one's experiences on general attitudes towards trust.[20] Were the results not robust to these additional controls, then the effect of the *IndvTrustOther* could be spurious and alternative accounts would be more plausible.

It is also interesting to note that the *AvgTrust* variable is positive and significant

[20] This doesn't mean they are from the same "culture," but the lack of a precise definition of culture in previous work prevents our making more precise statements. Additional analyses that broke apart respondents by race or religion revealed few differences across sub-groups though this is likely because of small sample sizes for many sub-groups. Furthermore, additional analyses that compared pro-social persons to those with individualistic and competitive orientations (van Lange *et al.*, 1997) did not change the avatar results.

in several specifications. This variable measures the average trustworthy score of the subject's evaluation of all avatars. In additional models (not reported here) that did not include the *AvgTrustOwn* and *AvgTrustOther* variables (they are of course highly correlated with *AvgTrust*), this variable is positive and significant in both the generated and evaluated models. This suggests additional support for Hypothesis 3. Individuals who perceive faces as being more trustworthy also appear to send more in the trust game. Of course, it is possible that when evaluating faces after the experiment individuals rationalized their evaluations conditional on the amounts they sent. However, it is unlikely that this drives the results. Between the game and evaluation stages, subjects went through a short break and had to fill out a set of demographic questions. Further, subjects went through many iterations of the trust game against many different opponents who were choosing the avatars. In the evaluation section, individuals were asked to simply evaluate the avatars along a scale. Thus, it is unlikely that a subject could recall a particular avatar and the amount they had sent in that part of the experiment, although this rationalization dynamic can not be ruled out completely.[21] Of course, by controlling for this variable the results of the key individual level measure, *IndvTrustOther*, is all the more interesting as its effect is identified off of deviations from average trust orientations.

Finally, I address Hypothesis 4 and compare sessions where avatars were randomly assigned to those where subjects chose their avatar. Hypothesis 4 suggests that any influence of the avatars will be eliminated when they are randomly assigned. Because Role 1 subjects know the assignment is random, any link between Role 2's intentions and their choice of avatar will be severed. In these sessions there was no communication between players, cheap or otherwise. In the final column of Table 3 I estimate a model using all the data along with a dummy variable for the sessions with generated faces used first. All models include a full set of interactions between each variable and an indicator variable for whether the session had avatars assigned randomly. Included but not reported are the control variables in Table 3 and their interactions.[22]

[21] Asking subjects to evaluate faces before the experiment would surely bias the results because it would prime them to think about trust.

[22] Before moving to the results, it is important to consider whether the post-experiment evaluations of

We see a negative interaction between *IndvTrustOther* and *Random*. While *IndvTrustOther* was positive and significant–indicating the relationship when avatars were chosen–the interaction, *IndvTrOthRandom*, was negative and statistically different from zero.[23] This suggests that in sessions where avatars were chosen, individuals were inferring something about the receiver's intentions based upon their choice of avatar. Put differently, the simple presence of a particular avatar on the screen was not necessarily consequential but instead the presence of the avatar given that it was chosen by an opponent is what was consequential for choice.

Because tobit is a non-linear model, substantive effects calculations are necessary to illustrate these interactive relationships. Figure 4 presents the results of a simulation that plots the predicted relationship between an individual's evaluation of the chosen avatar's trustworthiness using model *Int3*. This effect is calculated by shifting through $10th - 90th$ percentiles of the *IndvTrustOther* variable, conditional on being in sessions where avatars were randomly assigned versus chosen. In the simulation all other variables are set at their sample medians, though changing these values to other quantities does not influence interactive relationship plotted. The results show that in sessions where avatars were chosen, the amount sent increases with the perceived trustworthiness of the receiver whereas no such relationship exists for sessions where avatars are assigned.

4.3 Results Summary

Trustworthiness is a central theme in political life. Are perceptions of trustworthiness influenced related to physical features of the human face? The empirical results of the paper suggest such a connection. Consistent with Hypothesis 1, individuals more frequently chose avatars that were more trustworthy. This dynamic was particular to the trust game. Summary results from an additional experiment using an ultimatum game,

faces differed on average depending on whether someone had participated in an experiment with avatars chosen or randomly assigned. Difference-in-means tests showed no such significant differences in the generated and evaluated sets except for the TR1 face, which received a slightly higher average trustworthy evaluation in sessions where avatars were chosen (p=.07). No other differences were significant.

[23] This result holds under a range of specifications, including only using instances with frequently chosen avatars.

reported in the online appendix, show that dominant and higher threat/lower trust faces were chosen more frequently (see Figure 6).

There was support for Hypothesis 2 but only for the generated faces. There was more support for Hypothesis 3, that individual perceptions of face trustworthiness influence the amount sent. Controlling for a range of variables, individuals who perceived a face to be more trustworthy gave more than subjects that perceived their opponent's face to be less trustworthy. [24] Finally, it appears that individuals indeed were inferring something about trustworthiness from the *choice* of faces by the receiver in that these perceptions of trustworthiness had little effect when avatars were randomly assigned instead of chosen. Hypothesis 4 is supported and in this experiment "cheap talk" influenced choices. Consistent with some previous experimental work, non-costly signalling of intentions can influence behavior even when incentives are not aligned (Tingley and Walter, 2011). The influence of costless signalling in politics and economics is perhaps broader than standard models imply. Furthermore, these inferences were at least partially correct. Section 6.4 of the online appendix analyzes the amount returned: Role 2 players who chose more trustworthy avatars behaved on average in a more trustworthy way. The more they were sent the more they returned compared to individuals choosing less trustworthy avatars. While these experiments suggest an important role for physical features in influencing trustworthiness, which others note is central to political interactions, the experiments were nevertheless abstractions rather than linked to substantive political situations.

4.4 Follow-up Experiment: Mediator Choice

Next I report a short follow-up experiment where I embed the experiment in a more explicitly political context. As discussed throughout the introduction, the dynamics identified in this paper apply to a range of social contexts, including politics. Furthermore, trust is crucial in many areas of politics, as a range of authors point out (Levi and Stoker, 2000; Mishler and Rose, 1997; Hetherington, 1998; Putnam, 2000). Here I ask whether individuals would choose *mediators* for an international crisis situation where getting the

[24] Furthermore, in Section 6.3 of the online appendix I show that perceptions of trustworthiness explain more variation in the data compared to alternative dimensions of dominance and trust.

disputants to trust the mediator was crucial. Subjects selected from the same set of generated faces used above.

The experiment, fielded on Amazon's Mechanical Turk interface with 130 US based adults, began with the following prompt:

> "We would now like you to consider the following scenario. Try to think of yourself as if you were in the situation. You are the President of the United States. Recently there has been a serious conflict overseas involving two other countries. The US has decided to serve as a mediator between the disputing parties. It is crucial, above everything else, that these parties trust what the mediator says, and not seem like they will deceive the negotiators from the other countries. As President, you are able to choose who will be the mediator. You have been given a set of files with all equally qualified candidates, along with their pictures. The pictures are in the form of computerized renditions of their face. For each of the pictures below, report how likely would you be to choose each candidate."

The average likelihood of selecting each avatar is displayed in Figure 5. The results are largely consistent with the preceding experiment. Avatars that were selected more frequently in the behavioral experiment were more likely to be selected in the "mediator choice" simulation. However, the TW5 was only slightly more preferred than the TW3 face. Nevertheless, the orderings are correct.[25] While a number of more detailed experiments would be necessary to plot out how physical features influence choice of real world political agents, these results suggest the broad way that physical features social decision-making, including in politics.

5 Conclusion

I study a one-shot trust game where subjects see avatars that represent their partners. In some of the experiments subjects selected their avatars whereas in others the avatars

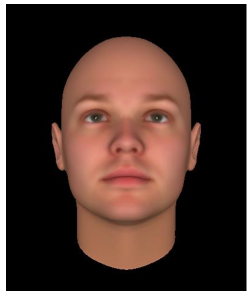
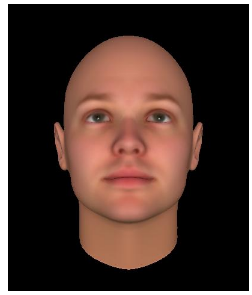
[25] The results hold if I exclude the handful of people who said they had at some point engaged in online tasks like "SecondLife" that feature avatars.

were randomly assigned. When subjects had the opportunity to choose avatars, they regularly chose avatars that ex ante are associated with higher levels of trustworthiness. Subjects rarely chose avatars that varied along the dominance dimension or were high in threat association. Despite knowing nothing about how the faces were generated, subjects intuitively gravitated towards more trustworthy faces. These faces have been argued to signal "approach", as opposed to avoidance, intentions. This provides new support for the approach taken by Todorov *et al.* (2008) and in other related work, but in an economically consequential environment. Furthermore, when Role 2 subjects (receivers) in the trust game chose more trustworthy looking faces, there was some evidence that they were sent more money by their Role 1 partners. This effect is strongest when an individual's perceived trustworthiness of the Role 2 avatar is used. These effects disappear when avatars were randomly assigned. This suggests two things. First, to some extent "cheap talk" was effective here, in that information about the intentions of the Role 2 person appear to have been communicated via avatar choice. Second, individuals who perceive faces as being more trustworthy also behave in a more trusting fashion. Hence individual variation in trustworthiness might also have to do with the processing of faces and not simply cultural or experiential variables.

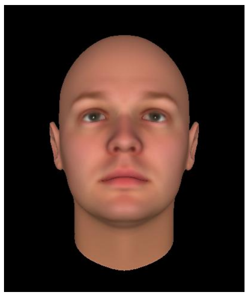
There are a number of implications for the study of politics and here I list only a few. First, most theories of candidate or agent choice suggest that non-physical attributes of candidates are relevant for choices. However, this work suggests that physical attributes of an individual may be relevant, but the character of these attributes could differ depending on the incentive situation. This conclusion parallels other work suggesting a link between facial characteristics and leader selection in times of war versus peace (Little *et al.*, 2007). Second, selecting agents that appear more trustworthy may be "cheap" and susceptible to imitation by others that are not trustworthy. But impact on behavior could still be consequential, at least in early interactions. On the margins political actors may wisely, if unconsciously, be selecting agents with physical first impression considerations in mind. In parallel work I am collecting neutral expression photos of every world leader since 1945 as well as samples of pictures from various diplomatic and military units of the United

States.

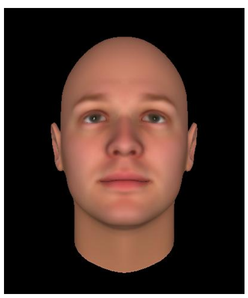
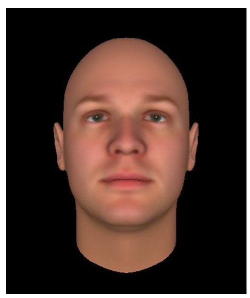
The results presented here prompt avenues for future work. One might take the approach in Eckel and Petrie (2011) and investigate how much individuals are willing to pay to be represented by different avatars. This would provide information on the perceived value of different facial characteristics. In ongoing experiments I investigate avatar choice and behavior in the ultimatum and power-to-take games (Bosman *et al.*, 2005), where I expect to find a greater role for threatening and dominant facial features. Similarly, little is understood about how the brain processes information contained in faces and translates these perceptions of other's intentions into economic and political choices. An open question is whether individuals who are less trusting of others also evaluate faces as being less trustworthy. Extending these studies to child subject pools (Antonakis and Dalgas, 2009) would help us test the role of inherited dispositions versus cultural learning. This and other work will help integrate emerging literatures in the social sciences on the role of appearance and social interaction. Finally, experiments on cheap talk might utilize both selected avatars but also permit more explicit forms of communication.

TW1 TW3 TW5

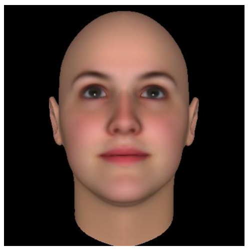


Dom1 Dom3 Dom5

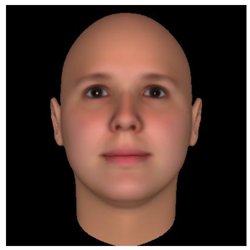




Threat1 Threat3 Threat5

Table 1: "Generated" pictures arranged by Todorov et al.'s dimensions of trust (TW), dominance (Dom), and threat (Threat). From left to right in each each dimension the face is -2 sd, 0 (mean), and +2 sd around the mean.

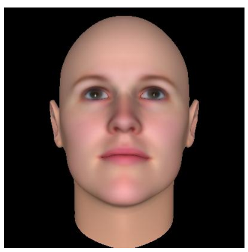


TR1

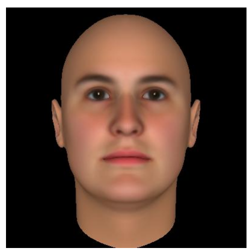


TR2

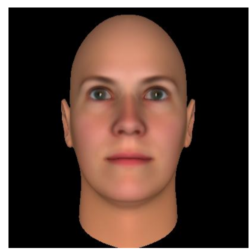
TR3



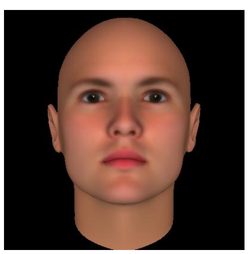
TR4



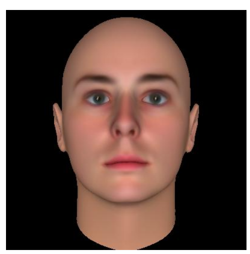
TR5



TR6



TR7



TR8

TR9

Table 2: "Evaluated" pictures arranged by Oosterhof and Todorov (2008) trust rankings, with TR1 being the most trustworthy and TR9 being the least trustworthy.

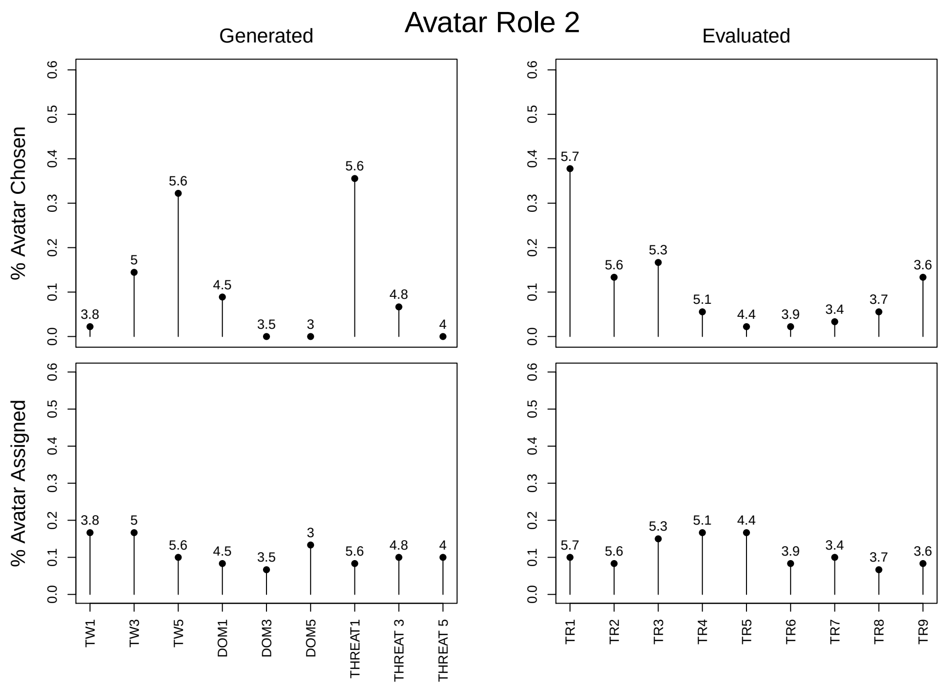


Figure 1: Percentage of time avatar avatar chosen by Role 2 person at the beginning of an experimental repetition. Top row for sessions where avatars were chosen ($N = 90$ in each) and bottom row for sessions were avatar was randomly assigned ($N = 60$ in each). Above each option is the average trustworthy score from the post-experiment survey.

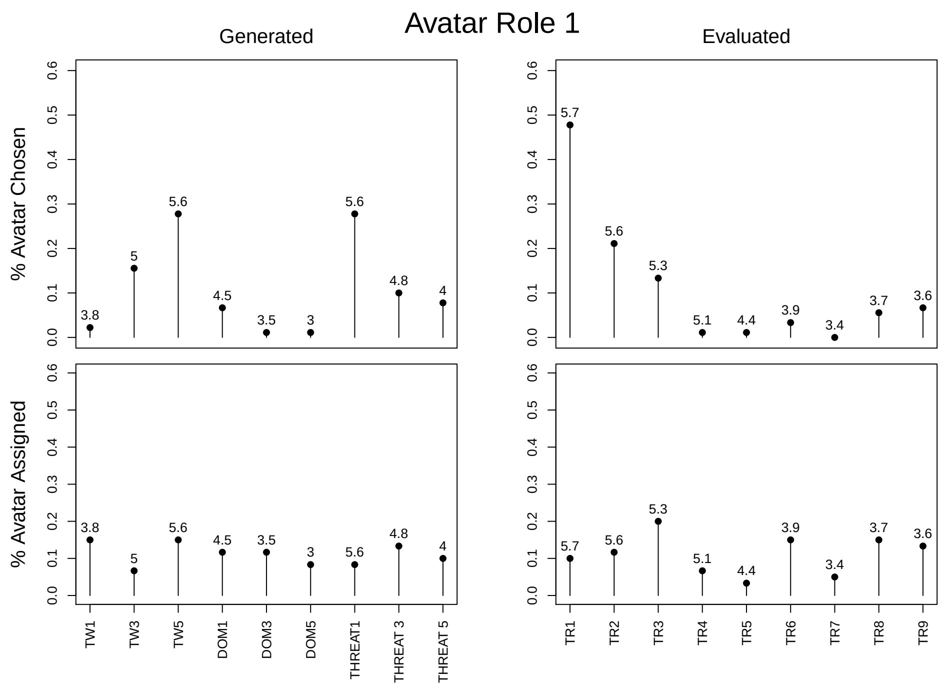


Figure 2: Percentage of time avatar avatar chosen by Role 1 person. Top row for sessions where avatars were chosen and bottom row for sessions were avatar was randomly assigned. Above each option is the average trustworthy score from the post-experiment survey.

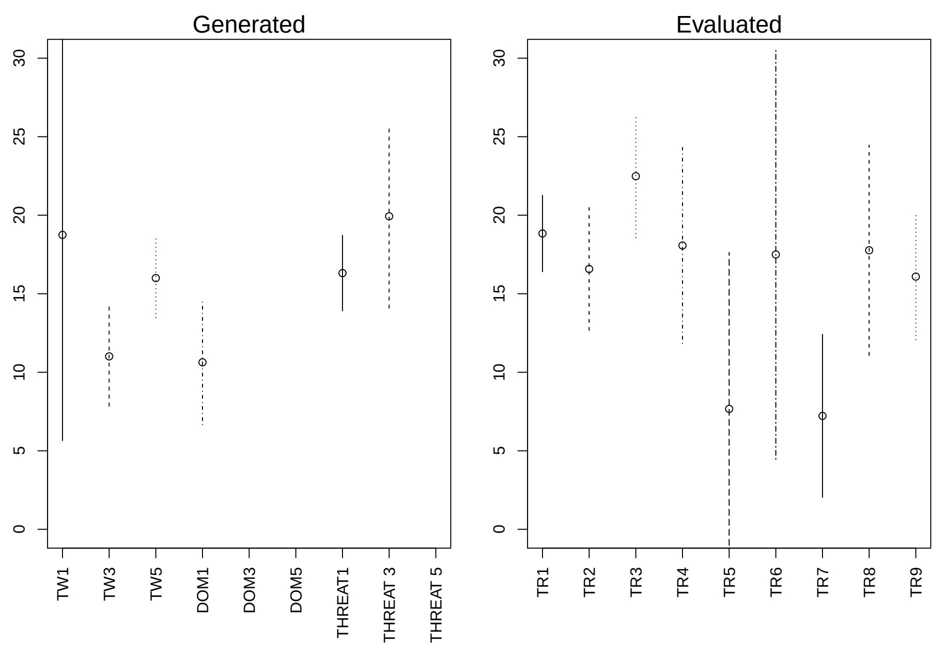


Figure 3: Distributions of amount sent by avatar chosen by Role 2 player. Means with 90% confidence intervals.

	All1	All2	All3	Gen1	Gen2	Gen3	Eval1	Eval2	Eval3	Random
model										
AvgTrustOth	1.71			5.59			-0.12			
	(1.79)			(3.69)			(2.10)			
IndvTrustOther		3.19*	2.61*		3.57+	3.74+		2.96*	1.60	4.37*
		(1.28)	(1.09)		(1.92)	(1.98)		(1.34)	(1.08)	(1.55)
IndvTrustOwn			2.72			-1.02			6.01+	
			(1.79)			(1.94)			(3.09)	
AvgTrust	10.26*	6.50+	3.74	8.46*	4.30	4.97	11.12*	7.67	-0.25	6.82
	(3.40)	(3.65)	(4.58)	(4.03)	(4.70)	(5.11)	(5.31)	(5.61)	(7.89)	(4.55)
Repetition	-6.01*	-6.29*	-6.09*	-8.64*	-8.16+	-8.14+	-8.71+	-8.79+	-8.14+	-7.47*
	(2.27)	(2.28)	(2.19)	(4.30)	(4.27)	(4.24)	(4.62)	(4.57)	(4.35)	(2.51)
WVSTrust	7.04+	6.30	5.51	4.90	4.65	4.83	9.19	8.21	6.05	7.72+
	(4.01)	(3.93)	(3.66)	(4.46)	(4.64)	(4.65)	(6.29)	(5.91)	(5.13)	(4.47)
Generated	-14.26*	-12.97*	-10.40+							-14.37*
	(6.35)	(6.30)	(5.76)							(7.08)
IndvTrOthRandom										-5.45*
										(2.73)
Random										-39.02
										(49.66)
Constant	-23.26	-13.92	-14.62	-40.99	-12.02	-10.78	-10.18	-9.80	-3.31	-28.07
	(21.69)	(19.04)	(19.34)	(34.15)	(21.22)	(21.09)	(36.29)	(34.86)	(35.45)	(23.57)
sigma										
Constant	34.26*	33.89*	33.72*	31.38*	31.19*	31.12*	36.59*	36.17*	35.22*	44.37*
	(4.47)	(4.31)	(4.20)	(3.90)	(3.81)	(3.85)	(6.64)	(6.45)	(6.08)	(5.04)
ll	-2272	-2264	-2258	-1078	-1075	-1075	-1190	-1185	-1172	-3671
BIC	4600	4582	4578	2199	2194	2199	2422	2412	2394	7444
N	924	924	924	462	462	462	462	462	462	1548

+ p<0.10, * p<0.05

Table 3: Tobit regression with amount sent dependent variable. Both generated and evaluated faces used. Robust standard errors clustered at the individual level in parentheses. Two-sided p-values reported.

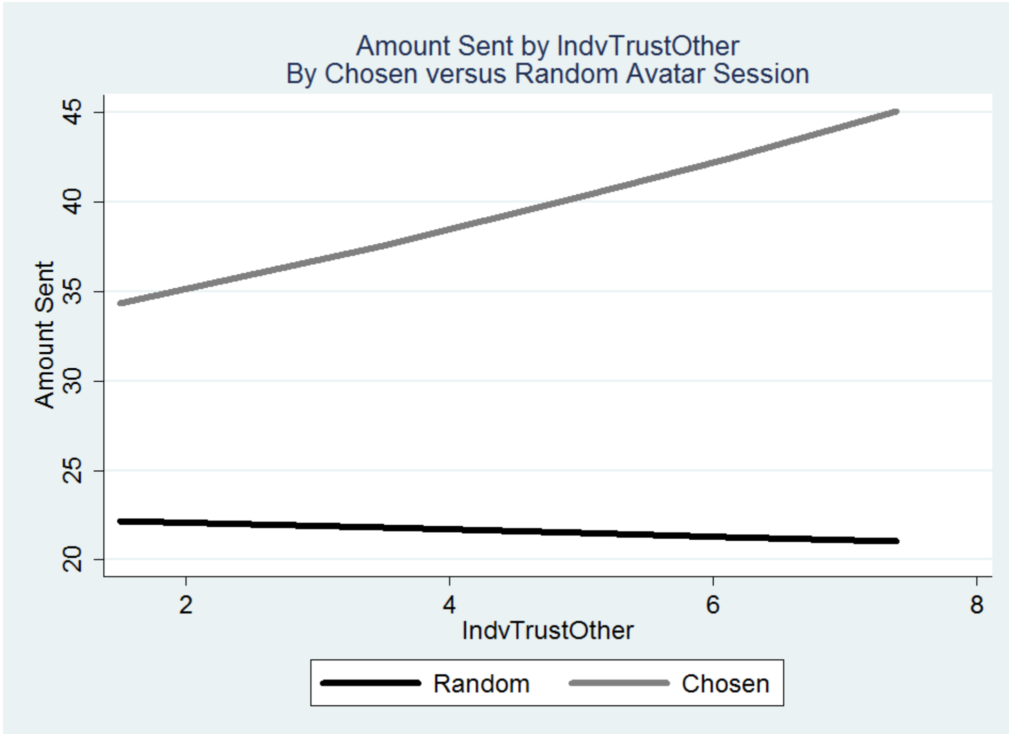


Figure 4: Amount sent is plotted against *IndvTrustOther* (individual level evaluation of face trustworthiness). The top line uses predictions based on avatars being chosen and the bottom line for sessions with avatars randomly assigned. Model estimated allowing for interaction between all covariates and random assignment. IndvTrustOther varied from sample 10th to 90th percentiles. All other variables set at their sample medians.

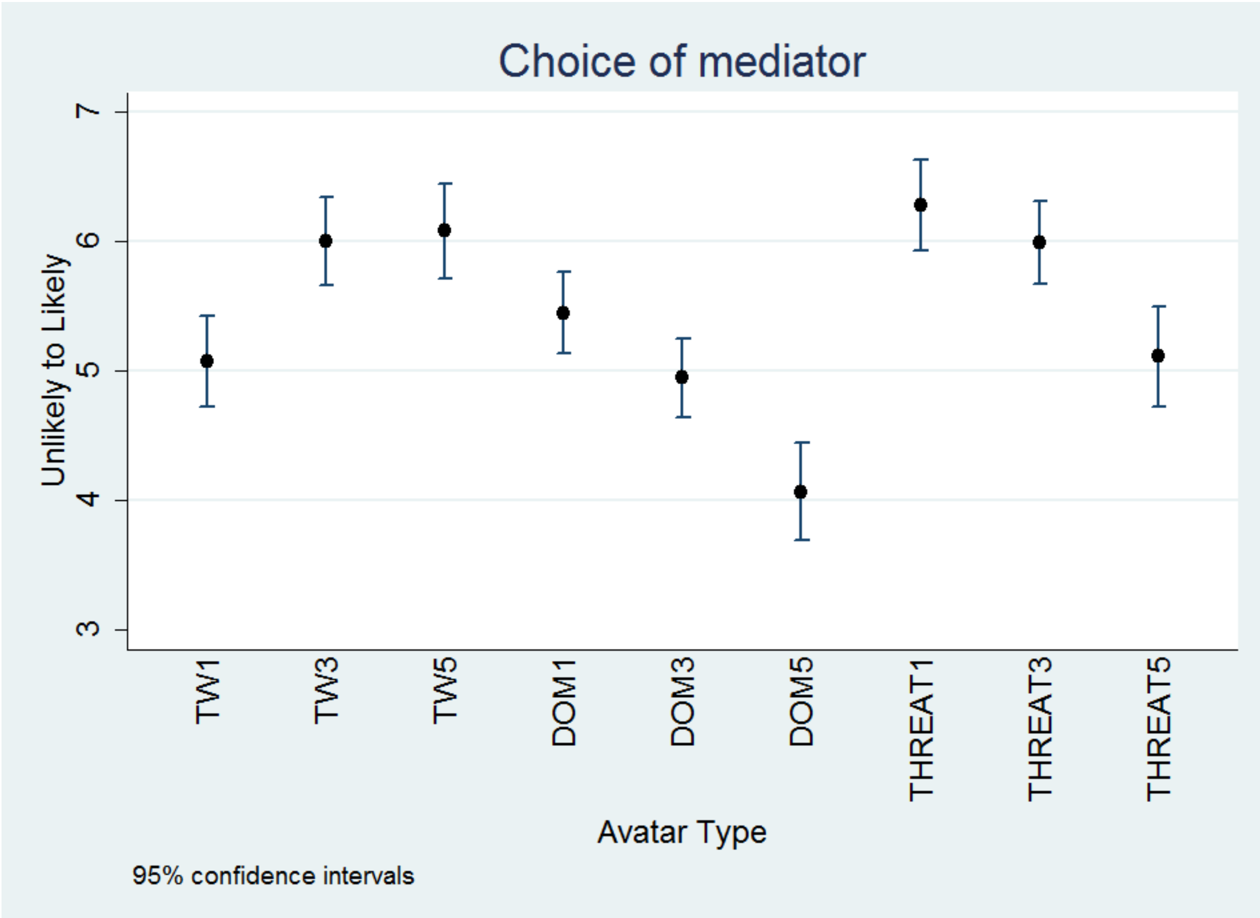


Figure 5: Likelihood would choose individual that looked like avatar to serve as mediator in international dispute. Higher values indicate a higher likelihood. Avatars generated with higher levels of trust were more likely to be chosen.

References

Abramowitz, A. (1989). Viability, electability, and candidate choice in a presidential primary election: A test of competing models. *The Journal of Politics*, **51**(04), 977–992.

Anderson, M. R. (February 2010). Community psychology, political efficacy, and trust. *Political Psychology*, **31**, 59–84(26).

Antonakis, J. and Dalgas, O. (2009). Predicting elections: Child's play! *Science*, **323**(5918), 1183.

Ashraf, N., Bohnet, I., and Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, **9**(3), 193–208.

Atkinson, M. D., Enos, R. D., and Hill, S. (2009). Candidate faces and election outcomes: Is the face-vote correlation caused by candidate selection? *Quarterly Journal of Political Science*, **4**(3), 229–249.

Austen-Smith, D. (1990). Information transmission in debate. *American Journal of Political Science*, **34**(1), pp. 124–152.

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

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, **10**(1), 122–142.

Biddle, J. and Hamermesh, D. (1998). Beauty, productivity, and discrimination: Lawyers' looks and lucre. *Journal of Labor Economics*, **16**(1), 172–201.

Bohnet, I., Greig, F., Herrmann, B., and Zeckhauser, R. (2008). Betrayal aversion: Evidence from brazil, china, oman, switzerland, turkey, and the united states. *American Economic Review*, **98**(1), 294–310.

Bosman, R., Sutter, M., and Van Winden, F. (2005). The impact of real effort and emotions in the power-to-take game. *Journal of Economic Psychology*, **26**(3), 407–429.

Bracht, J. and Feltovich, N. (2009). Whatever you say, your reputation precedes you: Observation and cheap talk in the trust game. *Journal of Public Economics*, **93**(9-10), 1036–1044.

Chiappe, D., Brown, A., and Dow, B. (2004). Cheaters are looked at longer and remembered in social exchange situations. *Evolutionary Psychology*, **2**, 108–120.

Crawford, V. (1998). A survey of experiments on communication via cheap talk. *Journal of Economic Theory*, **78**(2), 286 – 298.

DeBruine, L. M. (2002). Facial resemblance enhances trust. *Proceedings of the Royal Society of Biological Sciences*, **269**, 1307–1312.

Delgado, M. R. and Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, **8**(11), 1611–1618.

Eckel, C. C. and Petrie, R. (2011). Face Value. *American Economic Review*.

Engell, A. D., Haxby, J. V., and Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in human amygdala. *Journal of Cognitive Neuroscience*, **19**, 1508–1519.

Frank, R. (1988). *Passions Within Reason: The Strategic Role of the Emotions*. New York: W. W. Norton.

Glaeser, E., Laibson, D., Scheinkman, J., and Soutter, C. (1995). Measuring trust. *Quarterly Journal of Economics*, **115**(3), 811–846.

Haley, K. J. and Fessler, D. M. (2005). Nobody's watching? subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, **26**(245–256).

Hetherington, M. J. (1998). The political relevance of political trust. *American Political Science Review*, **92**(4), 791–808.

Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences*, **107**(25), 11163.

Kydd, A. (2007). *Trust and Mistrust in International Relations*. Princeton University Press.

Lawson, C., Lenz, G. S., Baker, A., and Myers, M. (2010). Looking like a winner: Candidate appearance and electoral success in new democracies. *World Politics*, **62**(04), 561–593.

Levi, M. and Stoker, L. (2000). Political trust and trustworthiness. *Annual Reviews of Political Science*, **3**(1), 475–507.

Little, A. C., Burriss, R. P., Jones, B. C., and Roberts, S. C. (2007). Facial appearance affects voting decisions. *Evolution and Human Behavior*, **28**(1), 18 – 27.

Mattes, K., Spezio, M., Kim, H., Todorov, A., Adolphs, R., and Alvarez, R. M. (2010). Predicting election outcomes from positive and negative trait assessments of candidate images. *Political Psychology*, **31**(1), 41–58.

Mazur, A., Mazur, J., and Keating, C. (1984). Military rank attainment of a west point class: Effects of cadets' physical features. *American Journal of Sociology*, **90**(1), 125–150.

Mishler, W. and Rose, R. (1997). Trust, distrust and skepticism: Popular evaluations of civil and political institutions in post-communist societies. *Journal of Politics*, **59**(2), 418–451.

North, M. S., Todorov, A., and Osherson, D. N. (2010). Inferring the preferences of others from spontaneous, low-emotional facial expressions. *Journal of Experimental Social Psychology*, **46**(1109–1113).

Oda, R. (1997). Biased face recognition in the prisoners dilemma game. *Evolution and Human Behavior*, **18**, 309–315.

Olivola, C. and Todorov, A. (2010a). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior*, **34**(2), 83–110.

Olivola, C. and Todorov, A. (2010b). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, **46**(2), 315–324.

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

Pinker, S. (2002). *The Blank Slate: The Modern Denial of Human Nature*. Viking.

Putnam, R. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton NJ: Princeton University Press.

Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster, New York.

Raftery, A. (1995). Bayesian model selection in social research (with discussion). *Sociological Methodology*, **25**, 111–196.

Rezlescu, C., Duchaine, B., Olivola, C., and Chater, N. (2012). Unfakeable Facial Configurations Affect Strategic Choices in Trust Games with or without Information about Past Behavior. *PLoS One*, **7**(3).

Said, C., Sebe, N., and Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, **9**(2), 260.

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

Stirrat, M. and Perrett, D. (2010). Valid Facial Cues to Cooperation and Trust. *Psychological Science*, **21**(3), 349.

Tingley, D. and Walter, B. (2011). Can cheap talk deter? an experimental analysis. *Journal of Conflict Resolution*, **55**, 994–1018.

Todorov, A. (2011). Evaluating faces on social dimensions. In A. Todorov, S. Fiske, and D. Prentice, editors, *Social Neuroscience: Toward Understanding the Underpinnings of the Social Mind*. Oxford University Press.

Todorov, A., Said, C. P., Engell, A. D., and Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, **12**(455–460).

Todorov, A., Pakrashi, M., and Oosterhof, N. N. (2010). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, **27**, 813–833.

van Lange, P., Otten, W., De Bruin, E., and Joireman, J. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology*, **73**(4), 733–746.

Van't Wout, M. and Sanfey, A. (2008). Friend or foe: the effect of implicit trustworthiness judgments in social decision-making. *Cognition*, **108**(3), 796–803.

Wentura, D., Rothermund, K., , and Bak, P. (2000). Automatic vigilance: The attention grabbing power of approach and avoidance related social information. *Journal of Personality and Social Psychology*, **78**, 1024–1037.

Willis, J. and Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, **17**(7), 592–598.

Winston, J., Strange, B., O'Doherty, J., and Dolan, R. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *nature neuroscience*, **5**(3), 277–283.

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

6 Online Appendix

6.1 Avatar choices in Ultimatum Game

In a separate set of experiments (with different subjects [$N = 68$] and analyzed separately) the experimental protocol was repeated but subjects played an ultimatum game instead of a trust game. In the ultimatum game Role 1 subjects were given 150 points. They would then propose a division of the points. Then Role 2 could decide whether or not to accept the division. Figure 6 presents avatar choices for Role 1 and Role 2 players in an ultimatum game. Role 1 players chose faces that were similar to faces chosen in the trust game, with high trust/low threat faces most frequently chosen. Role 2 choices were different. While the modal choice for generated faces was again the high trust TW5 and low threat THREAT1 faces, the frequency with which other faces were chosen was higher compared to Role 2 choices in the trust game (see Figure 1). In particular, the high threat and dominance faces were chosen with greater frequency. For evaluated faces, the least trustworthy avatar, TR9, was chosen as frequently as the TR1 avatar. This again contrasts with choices in the trust game.

6.2 Post-experiment survey

After completing the sixth repetition the paid portion of the experiment subjects took a short break and then filled out a computer-based post-experiment survey including demographic and personality questions. One component of this survey has already been described, where individuals scored all of the faces (which were randomly ordered) on dimensions of trust, dominance, and threat. Figure 7 presents results for the generated faces and Figure 8 for the evaluated faces. The results are as expected.

Finally, several sessions included a survey that also asked subjects to indicate how much they would send if the person in Role 2 chose a particular face.[26] While again these choices are from subjects after they participated in the trust game section, it represents an additional opportunity to see the influence of facial features, although this situation

[26] The first two sessions did not ask this question in the post-experiment survey.

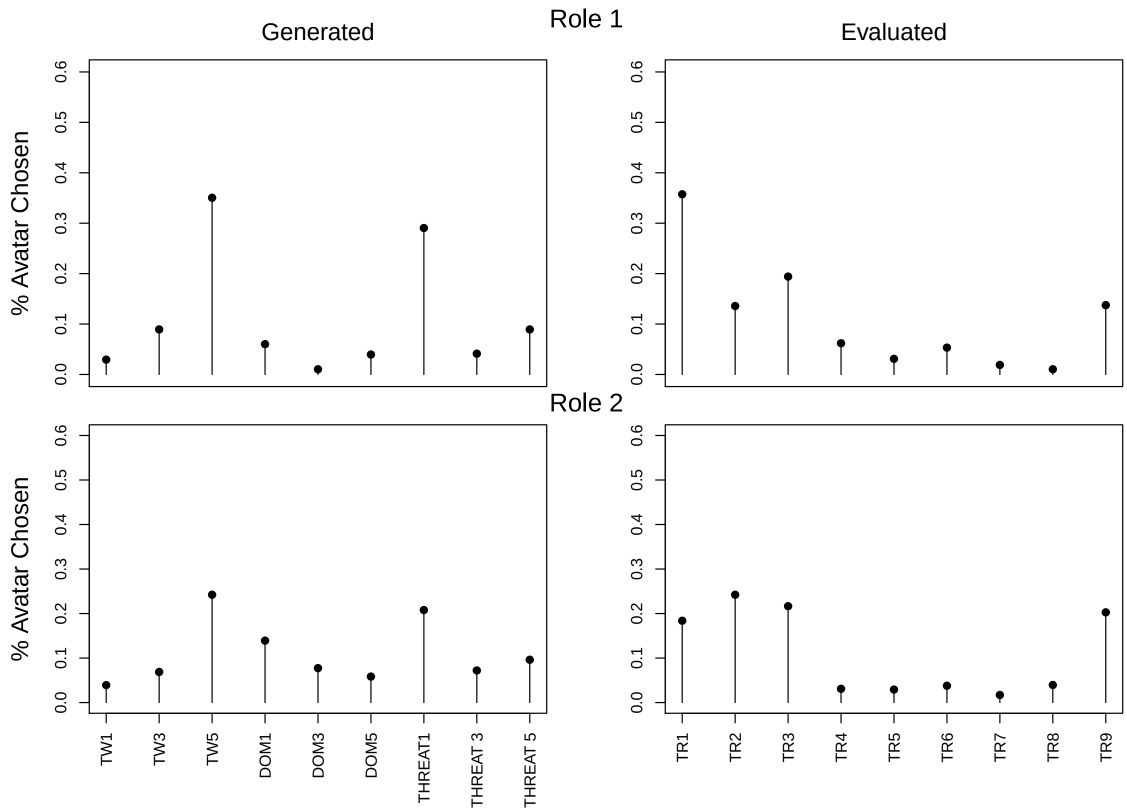


Figure 6: Avatar selection by Role 1 and Role 2 players in Ultimatum game. Experimental sessions used analagous protocols as in the trust game with three repetitions using generated faces and then three using evaluated faces.

is hypothetical. It is also helpful because, in the sessions with chosen faces, some avatars were chosen more frequently than others. Figure 9 provides violin plots for both the generated and evaluated faces. The plot on the left is for generated faces whereas the plot on the right is for evaluated faces.

The decline in amount that would hypothetically be sent is most apparent for the evaluated faces. For generated faces there was a decline in amount sent moving from the least dominant and least threatening (DOM1 and THREAT1) to the highest level of these characteristics. For the trust dimension faces, increases in the amount sent were smaller but still increasing from TR1 to TR3 as expected. These findings do not clearly support an expectation that changes in amount sent will matter most as we move along the trust dimension faces. Once again, we can turn to a multivariate analysis based upon the participants own perceptions of how trustworthy each of the different types of faces

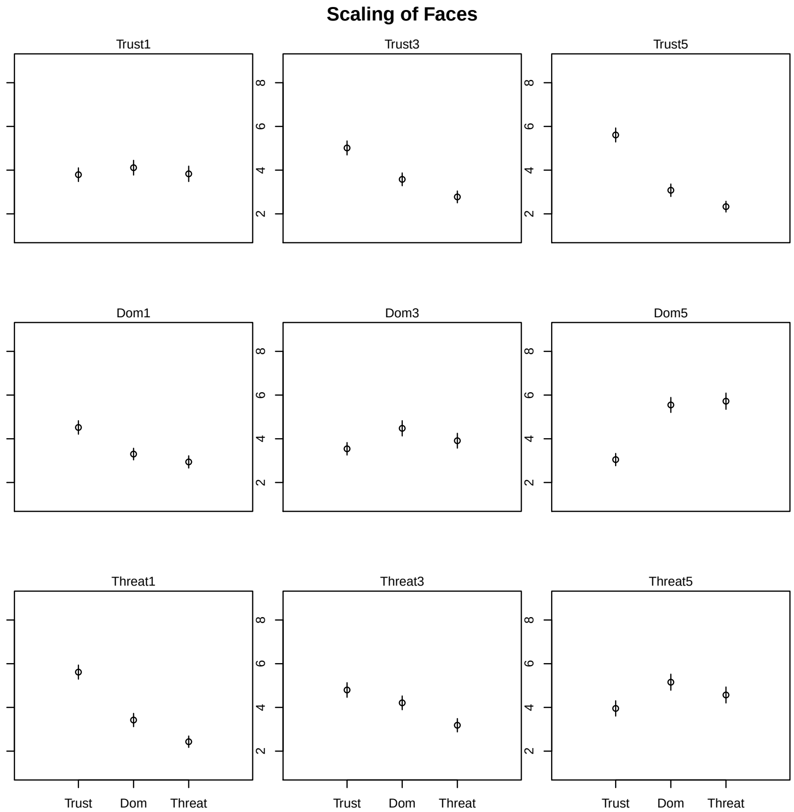


Figure 7: Post-experiment evaluations of generated faces along trust, dominance, and threat dimensions. Means with 95% confidence intervals.

look. Table 4 presents results for the generated faces and Table 5 presents results for the evaluated faces. An interesting pattern emerges. Looking first at the generated face results, the effect of trustworthiness evaluations have their strongest impact on faces that ex ante are less trustworthy or are more dominant and threatening. While individual variation in evaluations accounts for variations in stated amounts to be sent on faces with high threat/dominance and low trust, there is little variation explained by the evaluations for trustworthy or low dominance/threat faces. This result holds even if I do not control for participant demographics. A similar pattern is present with the evaluated faces. The effect of evaluations generally becomes stronger as we move from the most trustworthy TR1 face to the least trustworthy face TR9.

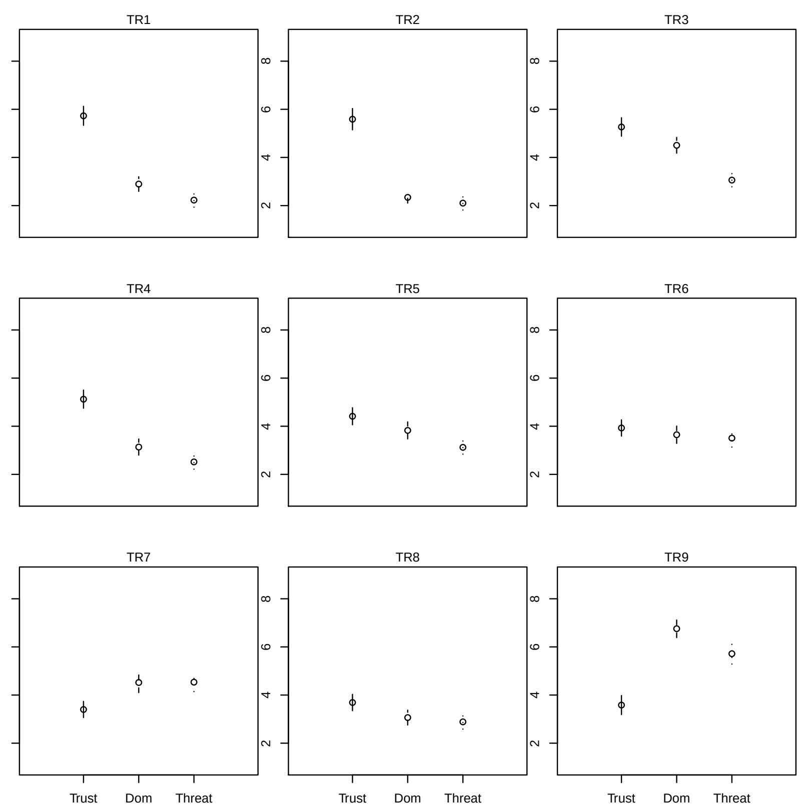


Figure 8: Post-experiment evaluations of evaluated faces along trust, dominance, and threat dimensions. Means with 95% confidence intervals.

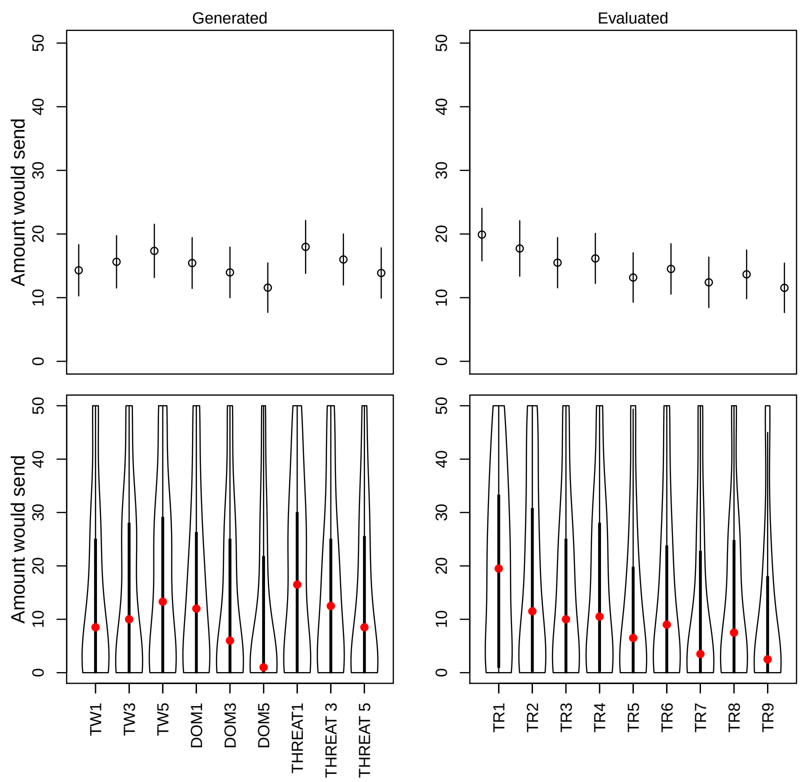


Figure 9: Post-experiment statement of amount the subject that would send were Role 2 to have selected that avatar. The plot on the left is for generated faces and the plot on the right is for evaluated faces.

	trust1	trust3	trust5	dom1	dom3	dom5	threat1	threat3	threat5
model									
trust1Trust	8.67*								
	(2.07)								
trust3Trust		3.36							
		(2.08)							
trust5Trust			0.75						
			(1.92)						
dom1Trust				3.18					
				(2.00)					
dom3Trust					7.66*				
					(2.17)				
dom5Trust						7.69*			
						(2.37)			
threat1Trust							1.71		
							(1.76)		
threat3Trust								4.96*	
								(1.75)	
threat5Trust									8.10*
									(1.68)
Male	-4.79	-8.38	-10.81+	-1.89	-2.52	-4.45	-8.96	-2.79	-1.87
	(5.96)	(6.70)	(6.14)	(6.05)	(6.07)	(6.69)	(6.06)	(5.81)	(5.70)
WVSTrust	0.73	2.88	3.12	3.79	2.44	3.25	1.68	4.32	3.90
	(3.51)	(3.66)	(3.37)	(3.30)	(3.47)	(3.83)	(3.42)	(3.38)	(3.26)
Constant	-22.28+	-9.08	7.53	-11.99	-23.31+	-27.34+	5.61	-23.09+	-32.85*
	(12.20)	(14.75)	(13.77)	(13.30)	(12.74)	(14.32)	(12.60)	(12.40)	(12.52)
sigma									
Constant	22.24*	25.44*	23.62*	23.24*	22.77*	24.33*	23.45*	22.17*	20.90*
	(2.87)	(3.27)	(2.80)	(2.84)	(2.95)	(3.28)	(2.79)	(2.74)	(2.71)
ll	-191	-203	-220	-211	-193	-180	-220	-204	-182
BIC	402	427	461	444	406	382	461	430	386
N	66	66	66	66	66	66	66	66	66

+ p¡0.10, * p¡0.05

Table 4: Tobit regression with hypothetical amount sent in post-experiment survey as dependent variable. For each generated face the the amount that would be hypothetically sent is regressed on the subject evaluations of trust. On all dimensions, faces furthest away from the positive evaluation show the strongest effect of the evaluation.

	TR1	TR2	TR3	TR4	TR5	TR6	TR7	TR8	TR9
model									
TR1Trust	2.76+								
	(1.55)								
TR2Trust		3.35*							
		(1.38)							
TR3Trust			4.57*						
			(1.48)						
TR4Trust				1.85					
				(1.54)					
TR5Trust					3.82*				
					(1.75)				
TR6Trust						4.58*			
						(1.83)			
TR7Trust							6.85*		
							(1.94)		
TR8Trust								5.43*	
								(1.75)	
TR9Trust									4.67*
									(1.71)
Male	-9.49	-6.57	-3.55	-10.82+	-5.02	-2.29	-3.17	-3.13	-1.74
	(5.96)	(6.29)	(5.66)	(5.83)	(6.00)	(5.88)	(6.02)	(5.56)	(6.58)
WVSTrust	3.24	2.19	1.30	1.77	1.98	2.00	-1.66	0.68	6.71+
	(3.35)	(3.52)	(3.19)	(3.31)	(3.47)	(3.42)	(3.49)	(3.17)	(3.78)
Constant	-1.83	-6.12	-13.54	4.81	-10.65	-11.56	-8.96	-9.86	-28.81*
	(12.08)	(12.12)	(11.48)	(10.97)	(11.98)	(11.80)	(11.29)	(10.81)	(13.75)
sigma									
Constant	22.67*	23.91*	21.67*	22.26*	22.79*	22.46*	22.60*	21.21*	24.06*
	(2.69)	(2.87)	(2.64)	(2.63)	(2.81)	(2.73)	(2.86)	(2.57)	(3.21)
ll	-218	-216	-208	-216	-205	-209	-196	-206	-186
BIC	457	453	436	453	432	438	413	433	393
N	66	66	66	66	66	66	66	66	66

+ p<0.10, * p¡0.05

Table 5: Tobit regression with hypothetical amount sent in post-experiment survey as dependent variable. For each evaluated face the amount that would be hypothetically sent is regressed on the subject evaluations of trust.

6.3 Performance of trust versus alternative measures

In the paper I look solely at how evaluations of avatar trustworthiness influence behavior. However, the research that this paper builds on also emphasizes two other important dimensions of the face: dominance and threat (which is conceptualized as a combination of dominance and trust (approach/avoidance)) (Todorov *et al.*, 2008). Todorov *et al.* (2008) leave as an outstanding question: "To what extent does the context of a decision affect the process of face evaluation"(pg. 459)? A related question is the extent to which evaluations of a face modulate decision-making in a particular decision context. The present study is uniquely situated to answer this question. While dominance and threat could indeed play important roles in other social interactions, the trust game isolates the role of trust and thus we should expect that measures of trustworthiness should matter more than measures of dominance or threat. Furthermore, because these authors treat "threat" as a combination of trust and dominance dimensions, we should expect that while evaluations of trustworthiness will play the strongest role, threat evaluations will play a slightly lesser role than trust, albeit in the opposite direction, and dominance to play no role at all. To explore this possibility, I estimate models that include the individual level analogues to the *IndTrustOwn* and *IndvTrustOther* variables for dominance (*IndDomOwn*, *IndvDomOther*) and threat (*IndThreatOwn*, *IndvThreatOther*) dimensions and report the results in Table 6. Models M1-M5 include only evaluations of the Role 2 avatar. The first two models include all three evaluations together, while the next three models enter the variables separately. Evaluations of trust consistently have the strongest effect in these models. Not only is the magnitude of the slope coefficient strongest for the *IndvTrustOther* variable, but the strongest model fit as reflected by the smaller Bayesian Information Criterion (BIC, (Raftery, 1995)) is the model with *IndvTrustOther* and no other face evaluation variable.[27] All of these results hold whether or not the subject level *Male*, *WVSTrust*, and *AvgTrust* are included.

[27] Furthermore, this model also improves model fit compared to models with only the session and individual covariates.

	m1	m2	m3	m4	m5
model					
IndvTrustOther	5.23*	2.71+	3.18*		
	(1.43)	(1.38)	(1.30)		
IndvDomOther	2.71*	1.76		-0.35	
	(1.27)	(1.21)		(0.93)	
IndvThreatOther	-1.81	-2.15			-2.21+
	(1.67)	(1.51)			(1.15)
Male		-1.11	-1.87	-2.48	-2.07
		(6.16)	(6.10)	(6.17)	(6.06)
WVSTrust		6.30	6.46+	7.22+	6.78+
		(3.89)	(3.92)	(3.97)	(3.96)
Repetition	-6.57*	-6.32*	-6.27*	-6.00*	-6.14*
	(2.24)	(2.31)	(2.29)	(2.29)	(2.33)
Generated	-12.93*	-13.18*	-12.76*	-12.54+	-13.05*
	(6.31)	(6.43)	(6.41)	(6.58)	(6.63)
Reverse	-25.22*	-25.65*	-25.86*	-27.24*	-27.19*
	(7.41)	(6.50)	(6.49)	(6.62)	(6.52)
AvgTrust		6.37+	6.32+	10.20*	10.55*
		(3.65)	(3.65)	(3.47)	(3.41)
Constant	14.66	-10.86	-13.00	-13.42	-8.14
	(12.65)	(19.92)	(19.55)	(19.64)	(19.84)
sigma					
Constant	34.60*	33.77*	33.91*	34.29*	34.09*
	(4.31)	(4.26)	(4.30)	(4.46)	(4.40)
ll	-2279	-2261	-2264	-2272	-2268
BIC	4613	4597	4589	4606	4597
N	924	924	924	924	924

+ p<0.10, * p<0.05

Table 6: Tobit regression with amount sent dependent variable. Models compare the role of trust, dominance, and threat evaluations. Robust standard errors clustered at the individual level in parentheses.

6.4 Amount Returned

Table 7 estimates the amount returned to Role 1 as a function of the amount sent, the average and individual level avatar trust measures, and a set of control variables. In the paper we saw that Role 1 players sent more to Role 2 players who chose a more trustworthy avatar. Here I evaluate whether this cheap talk signal was reciprocated. In particular, are individuals who chose more trustworthy avatars also likely to return a higher amount, conditional on the amount sent? To investigate this I estimate models with an interaction between the amount sent and the avatar trust measures. *AvgTrustOwnSent* is an interaction between the average trust evaluation of Role 2's avatar and the amount sent to Role 2 by the Role 1 player. *AvgTrustOtherSent* is an analogous measure, but uses the average trust score of the Role 1 player. Because the amount returned is from the perspective of the Role 2 player, "own" and "other" refer to the Role 2 and Role 1 avatars respectively. Similarly, *IndvTrustOwnSent* is the interaction between Role 2's own evaluation of the trustworthiness of their avatar and the amount sent. If Role 2 players were correctly signal their trustworthy intentions, we should see a positive coefficient on the *AvgTrustOwnSent* and *IndvTrustOwnSent* variables. Furthermore, because Role 2's decision is at the terminal node of the game, the trustworthiness of the Role 1 avatar might be expected to be irrelevant and hence not significantly different from 0.

These hypothesized patterns are largely present in the data. The coefficient on *AvgTrustOwnSent* is positive and significant. Role 2 subjects that chose more trustworthy avatars also returned more the more they were sent. The coefficient on *AvgTrustOtherSent* was not significantly different from 0. The coefficient on *IndvTrustOwnSent* is also positive and either significant or very close to significant with the two-tailed p-values. Finally, models R1 and R2 replicate models M3 and M6 but are estimated only with sessions where avatars were randomly assigned. The positive and significant interactions on the *AvgTrustOwnSent* and *AvgTrustOtherSent* variables disappear. Plotting the substantive effects support these interpretations. However, instead of presenting those figures I conduct a test analogous to the first test in Section 4.2 that compared the amount sent by whether the Role 2 player chose the more trustworthy TW3 avatar versus the slightly less trustworthy TW2 avatar (in repetitions using the Generated faces). There we saw that on average individuals sent more to people who chose the TW3 avatar. Here I investigate whether there is an interaction between the choice of TW2 versus TW3 and the amount sent. Based on a tobit model including the same controls as in Table 7, Figure 10 plots the amount returned as a function of the amount sent by whether an individual chose the TW3 or TW2 avatar. The amount returned by those with the TW3 avatar is nearly always higher compared to those with the TW2 avatar, and this distance is increasing in the amount sent.

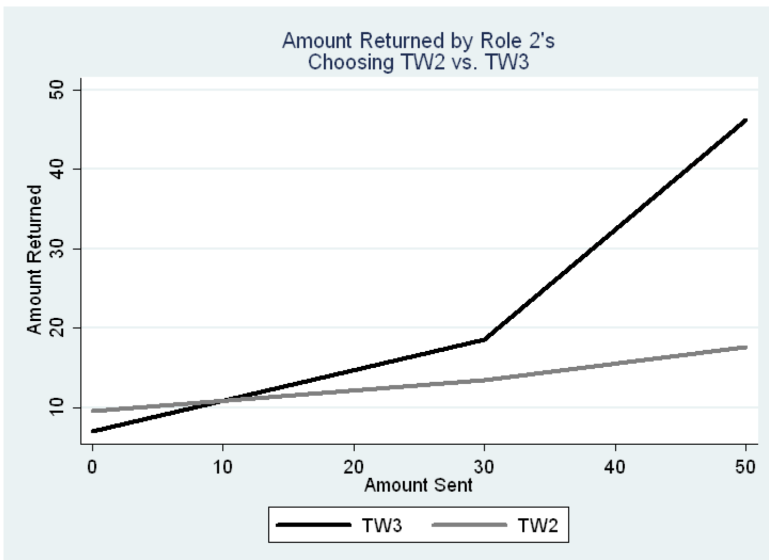


Figure 10: Amount returned as a function of amount sent and Role 2 avatar choice.

model	M1	M2	M3	M4	M5	M6	R1	R2
amountsent	0.33	1.49*	0.12	1.28*	1.85*	1.45*	2.68*	1.11*
	(0.71)	(0.63)	(0.82)	(0.43)	(0.48)	(0.55)	(0.84)	(0.39)
AvgTrustOwn	-6.32	2.09	-6.17				2.08	
	(4.14)	(3.80)	(4.20)				(4.49)	
AvgTrustOwnSent	0.30*		0.29*				-0.19	
	(0.13)		(0.13)				(0.13)	
AvgTrustOth	3.63	1.63	2.35				-1.87	
	(2.34)	(3.72)	(3.76)				(3.58)	
AvgTrustOthSent		0.09	0.05				0.05	
		(0.13)	(0.13)				(0.10)	
IndvTrustOwn				-0.89	2.43	-1.69		-0.18
				(2.25)	(1.58)	(2.12)		(1.78)
IndvTrustOwnSent				0.11		0.14*		0.06
				(0.07)		(0.07)		(0.06)
IndvTrustOther				1.88	1.63	3.53+		-5.79*
				(1.84)	(2.09)	(1.92)		(2.14)
IndvTrustOtherSent					0.01	-0.06		0.14*
					(0.08)	(0.08)		(0.06)
Male	-8.08	-8.58	-8.09	-8.09	-8.27	-8.43	-21.99+	-22.79+
	(8.22)	(8.40)	(8.21)	(8.04)	(7.99)	(8.07)	(11.58)	(12.21)
WVSTrust	0.48	0.43	0.49	-0.60	-0.44	-0.54	9.28+	8.71
	(4.75)	(4.88)	(4.75)	(4.66)	(4.76)	(4.66)	(5.42)	(5.70)
AvgTrust	5.83	5.62	5.80	-0.30	0.17	-0.25	16.57*	14.97*
	(4.83)	(5.06)	(4.84)	(5.13)	(5.34)	(5.12)	(5.31)	(5.04)
Generated	-3.40	-4.22	-3.52	1.35	0.81	1.21	4.21	4.11
	(4.60)	(4.42)	(4.54)	(4.17)	(4.19)	(4.23)	(5.25)	(5.12)
Repetition	-2.30	-2.35	-2.28	-2.62+	-2.52+	-2.57+	-4.63*	-4.93*
	(1.52)	(1.50)	(1.51)	(1.48)	(1.47)	(1.48)	(1.58)	(1.59)
Reverse	-13.06+	-12.56+	-12.89+	-12.43	-12.54	-12.17	-8.32	-7.74
	(7.62)	(7.58)	(7.55)	(7.72)	(7.65)	(7.64)	(10.22)	(10.05)
Constant	-48.49	-82.24*	-42.94	-40.91*	-60.95*	-46.05*	-135.59*	-98.60*
	(34.55)	(39.21)	(37.04)	(18.97)	(22.61)	(19.11)	(46.67)	(26.73)
sigma								
Constant	32.08*	32.47*	32.07*	31.79*	32.18*	31.80*	34.39*	34.28*
	(3.37)	(3.36)	(3.37)	(3.44)	(3.36)	(3.44)	(4.21)	(4.15)
ll	-1441	-1445	-1441	-1438	-1441	-1437	-1317	-1314
BIC	2964	2972	2970	2957	2964	2963	2718	2712
N	924	924	924	924	924	924	624	624

+ $p<0.10$, * $p<0.05$

Table 7: Tobit regression with Role 2's choice of amount returned as dependent variable. Interactions between amount sent and avatar trust ratings included.