

Lund Psychological Reports

Volume 16, No. 1, 2017

**Do we trust the poor?
Probing a game
paradigm for measuring
discriminatory behavior**

Anna Lindqvist and Fredrik Björklund

Department of Psychology, Lund University, Sweden



LUND
UNIVERSITY

Lund Psychological Reports

Editor: Magnus Lindgren

ISSN 1404-8035

Abstract

The aim of the present research was to evaluate a computerized version of the *Trust game* as a method for studying real-life decisions and discriminatory behavior towards the poor. The stereotype of the poor typically consists of low perceived competence and medium/high perceived warmth. According to the BIAS map (Behaviors from Intergroup Affect and Stereotypes), this should trigger behavioral tendencies of passive harm towards the poor, as compared to the ‘average man’ which served as our control condition, and could be expected to receive facilitatory behavior. The participants ($n = 210$) played a trust game online with someone portrayed to have either low or medium-high socio-economic status – while they in fact played with the computer. Their task was to decide to what extent they could trust the other player: if both collaborated they would receive the maximum reward. The results indicate that very few participants suspected that they were not playing with an actual person. Some of their responses in the trust game can be interpreted as being in line with the BIAS map, i.e. tendencies of passive harm towards the player with low SES and passive facilitation towards the player with medium-high SES. This suggests that a computerized trust game with a simulated other player may be a viable method for investigating real-life collaborative behavior. Remaining psychometric challenges are discussed.

Keywords: stereotype content model, BIAS map, trust game, game theory, poor, collaboration

Do we trust the poor? Probing a game paradigm for measuring discriminatory behavior

Anna Lindqvist and Fredrik Björklund

Department of Psychology, Lund University, Sweden

Stereotypes are beliefs about the attributes characterizing a social category or a group of people (Ashmore & Del Boca, 1981). They are sometimes referred to as the cognitive aspect of intergroup bias, where the other components are affective (emotional prejudice) and behavior (in terms of discrimination; Esses & Dovidio, 2002; Fiske, 1998). Stereotypes are assumed to influence behaviors towards individual members of the stereotyped group. Previous research reveals that the content of the stereotype influences whether the behavior can be expected to be prosocial, antisocial or neutral (Cuddy, Fiske, & Glick, 2007). However, the majority of studies so far have focused on the perception of groups rather than on individual group members, and concerned judgments/ratings rather than actual behavior towards members of these groups.

To estimate the importance of stereotypes for real-life behavior, researchers need to turn to other methods than the standard questionnaire studies and laboratory, or vignette, experiments. The primary aim of this report is to describe a promising method for doing this without compromising too much with either internal validity (e.g., experimental control) or external validity (the sense of interacting with an actual person). We investigate behavior towards poor individuals, by testing how the stereotype about individuals with low socio-economic status (SES) affects people's choices in an economic game-paradigm. More specifically, we investigate how the stereotype of the poor affects people's willingness to trust someone with low SES.

The stereotype of the poor

Stereotype content can be organized into the two dimensions of warmth and competence (*Stereotype Content Model*, SCM; Fiske, Cuddy, Glick, & Xu, 2002). The warmth dimension originates in the desire to determine others' intentions: Is this person sincere, friendly and trustworthy? The competence dimension has its origin in the desire to determine others' capability to pursue their intentions: Is this person skillful, creative and confident? The warmth dimension of SCM includes traits such as good-natured and trustworthy, while the

competence dimension includes traits such as capable and intelligent. These two dimensions are argued to be universal and central in both interpersonal and intergroup perception (Cuddy et al., 2007; Fiske, Cuddy, & Glick, 2007; Fiske et al., 2002; Fiske, Xu, Cuddy, & Glick, 1999). As identified by Fiske and colleagues (2002) in a US sample, the stereotypes may range from low warmth/low competence (LW/LC: welfare recipients) to medium warmth/medium competence (MW/MC: gay men) to high warmth/high competence (HW/HC: Christians) with all kinds of variations in between those extremes (e.g., LW/MC: Hispanics; LW/HC: rich).

According to Cuddy and colleagues (2007), people are more interested in identifying warmth-related traits than competence-related traits (see also Wojciszke, Bazinska, & Jaworski, 1998). The warmth dimension is also evaluated more quickly than the competence dimension (Willis & Todorov, 2006), probably because low-warmth traits signal more potential threat than low-competence traits do (Cuddy et al., 2007; Reeder, 1993; Wojciszke, 2005). The current report examines discrimination in relation to the warmth dimension, in terms of cooperation and trust. We base our predictions on the BIAS map (*Behaviors from Intergroup Affect and Stereotypes*; Cuddy et al., 2007), which describes how stereotypes of different content activate four different intergroup behavioral tendencies: Active harm (LW/MC), passive harm (MW/LC), active facilitation (HW/MC), passive facilitation (MW/HC). According to the BIAS map, stereotypes with differences in the SCM ratings will result in different kinds of discrimination. We have previously shown that the stereotype of the poor constitutes of MW/LC (Lindqvist, Björklund, & Bäckström, 2016), and accordingly predicted that individuals defined as poor will be met by the behavioral tendency of passive harm, being discriminated in relation to the competence dimension, but less so in relation to the warmth dimension. Passive harm is not explicit, but still has consequences for individuals of the stereotyped group (Cuddy et al., 2007). It might in this case, for example, manifest behaviorally by neglecting the welfare of a poor individual, or by not providing service to hir¹, while the attitude expressions towards the poor might vary between pity (related to slightly higher warmth) and contempt (related to slightly lower warmth).

Operationalization of cooperation

In previous research on the stereotype of the poor and its interpersonal consequences, the main approach has been to use vignettes where participants read about different

¹Hir refers to an individual of any gender.

individuals whose socio-economic status (SES) is manipulated (e.g., as either low or medium-high, as in Lindqvist & Björklund, 2017). In the current research, we wish to increase the external validity by testing actual behavior rather than judgments of hypothetical scenarios. One method for testing the impact of stereotypes in real-life situations related to trust, is to let the participants play financial cooperation games with individuals representing the target stereotypes (or a contrasting stereotype), and compare the results. If this method is successful, the outcome of the present research should map onto the predicted tendencies of passive harm towards the poor, described by Cuddy and colleagues (2007).

Moving from hypothetical situations in the lab to real situations with actual consequences for the participants has been a successful approach for testing real-life behavior in previous research. Other researchers have, for example, found that participants' behavior differ between hypothetical versus actual financial games (i.e. where the participants are playing with real money; Rydval & Ortmann, 2004). For example, people are less generous when playing with real money, because there is more at stake when the situation is not fictive (Forsythe, Horowitz, Savin, & Sefton, 1994; Sefton, 1992). A real game, rather than a hypothetical, should hence provide a more valid estimate of how stereotypes affect actual interpersonal behavior. In the current research, we use a *Trust Game* for investigating cooperative behavior, since we want to analyze the warmth aspect of the stereotype.

Computerized games simulating real players. One important aspect of the game design is to make the participants believe that they play with real persons, while they in fact play with a computer. The advantage of using a computer is increased control. Although it is not uncommon to use computerized cooperation games simulating another real player (as in e.g., Belli, Rogers, & Lau, 2012; Smith et al., 2009; Wilson & Eckel, 2006), the method can be quite risky. Participants tend to cooperate less if they understand that the other player is simulated (see e.g., Bottom, Holloway, Miller, Mislin, & Whitford, 2006; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). This may be due to the sense of nobody being “harmed” if the participants do not cooperate. To be able to claim validity in a computerized trust game, it is therefore crucial to check whether the participants believed that they were interacting with a person or not (see Johnson & Mislin, 2011 for a discussion). A simple but efficient method is to ask the participants if they noticed something about the other player, as Belli and colleagues did (2012).

Trust game

The original version of the trust game, as used in experimental social sciences, was developed by Berg, Dickhaut and McCabe (1995). In this game, anonymous players were paired together in a lab. Player 1 (sender) received \$10 and could keep these, or send any amount to Player 2 (returner). The sender put the money to be sent in an envelope, and the returner received three times this amount ($f=3$ is the most common factor). The returner then decided how much or little to send back to the sender.

According to game theory, since the returner has no incentive to return any of the money, the *Nash equilibrium* is for the sender to keep the \$10, sending none to the returner (see e.g., Berg, Dickhaut, & McCabe, 1995; Cardenas, Chong, Ñopo, Horowitz, & Lederman, 2009; Johnson & Mislin, 2008). Nevertheless, the senders tend to send money to the returners. Berg and colleagues (1995) found that participants playing the role of sender on average sent \$5.16 of their \$10 to the returner.

How stereotypes affect decisions in the trust game. Most research on how people make decisions in the trust game has focused on strategies in repeated games (as in e.g., Anderhub, Engelmann, & Güth, 2002; Engle-Warnick & Slonim, 2006), or on demographics of the participants that might explain how they choose to play in the game, such as gender (Croson & Buchan, 1999) and age (Belli et al., 2012). In the present report, we use a one-shot game, where the participants only play one round of the game. This strategy is used to avoid that the participants gain any information on the other player's game tendencies, such as previous decisions in the game, which most likely would influence their own behavior in the game. Instead, we strive to create a situation where participants may base their decisions on the limited information we provide about the other player – where we activate the target stereotype.

Most versions of the trust game have had anonymous players (see Johnson & Mislin, 2011 for a summary). It is relatively unusual to analyze how the participants are affected by stereotypes about the other player, but a couple of earlier studies can be found: Carlin and Love (2013) showed that stereotypes of political party affiliations affect the amount the sender sends to the returner, where an ingroup-favoring tendency could be shown. Wilson and Eckel (2006) showed that attractive players are trusted more. Cañadas and colleagues (2015) did *not* identify any differences in behavior, depending on the other player's ethnicity. However, the participants in their study knew that they were playing with bots, which decreases the validity of the game (cf., the discussion by Johnson & Mislin, 2011). Their results might, for example, be affected by social desirability.

Very little previous research has concerned the role of socio-economic status. Only one study has analyzed how the returners wealth affected the amount sent by the sender (Brülhart & Usunier, 2012). The hypothesis was that senders would have altruistic motives and send larger amounts to poor returners, which the researchers did not find any evidence of.

Study 1: Pilot Study

In the first study, which concerned the features of the stimulus person, we wanted to validate the information of the other player that we were planning to use as manipulation in the trust game. The aim was to create one version which activates the stereotype of the poor (i.e., the target stereotype), and one version that could be used as control by activating ‘an ordinary guy’.

Stimulus material and procedure

We created a short description with two versions of a middle-aged man, David (see Table 1), where his socio-economic status (SES) was manipulated in terms of occupation and his self-rated SES (from 1-10, where 1 = those worst off). The information about David was written in a bullet list, to make it easy for the participants to take in the information, and was presented in an online questionnaire.

The participants were randomized to read the low SES version (SES = 2), or the medium-high SES version (SES = 6), and were then asked to indicate how they perceived David in terms of warmth (friendly, well-intentioned, trustworthy, warm, good-natured, sincere) and competence (competent, confident, capable, efficient, intelligent, skillful). Ratings of the twelve items were made on a 5-point Likert scale where higher values imply higher warmth and higher competence, respectively (cf., SCM; Fiske et al., 2002). They were also asked to share their impression of David as a free-text response. The online questionnaire took about 5-10 minutes to complete.

Participants

The 68 participants were between 19-58 years ($M = 34.00$, $SD = 9.46$). They were self-defined as 35 women (51.5 %), 32 men (47.1 %) and 1 with non-binary gender-identity (1.5 %). All participants were from the United States, and were recruited through the web platform *MTurk* (www.mturk.com).

Table 1. The information about the other player used as stimulus material in the trust game, here divided into the two different versions/manipulations of low SES and medium-high SES. Warmth and competence show mean values of ratings from Study 1, from 1-5 where higher values imply more warmth/competence, standard deviation within parentheses.

version	name	age	country	occupation	SES	warmth	competence
low SES	David	43	USA	unemployed	2	3.11 (.82)	2.51 (.70)
medium-high SES	David	43	USA	full-time employed	6	3.62 (.56)	3.77 (.50)

Results

SCM ratings. These ratings of warmth and competence corresponded well to earlier SCM ratings of these two target groups: There was a significant difference both in rated warmth and in rated competence between the low SES version and medium-high SES version, $t_{warmth}(66) = 2.29$, $p = .005$, $d = .72$, $t_{competence}(66) = 8.44$, $p < .001$, $d = 2.05$. As revealed in Table 1, showing the mean values of the ratings, the low SES version of David received low-medium warmth/low competence ratings, while the medium-high SES version received medium-high warmth/medium-high competence ratings.

Free-text responses. We also wanted to validate the stereotype content by analyzing the associations the participants made when reading the two versions of the manipulation text. All words/descriptors for each version were counted, where synonyms (such as trustworthy–reliable–honest; man–male–guy; average–typical–ordinary) were aggregated. In total, the low SES version resulted in 18 different descriptors, while the medium-high SES version resulted in 29 different descriptors. All descriptors are presented in word clouds, where the size reflects the frequency of the word, see Figure 1 (low SES) and Figure 2 (medium-high SES).

As can be seen in Figure 1, the most common descriptor of the low SES version of David was “poor”. He was also described as unemployed, with no money. When having a SES of 6, the most prominent description of David was “average”. He was also described as being a hard-worker, and a middle-aged man. Being a middle-aged man was also relatively common descriptors in the SES=2 version. The medium-high SES version of David was also described as being middle-class, as well as being reliable.

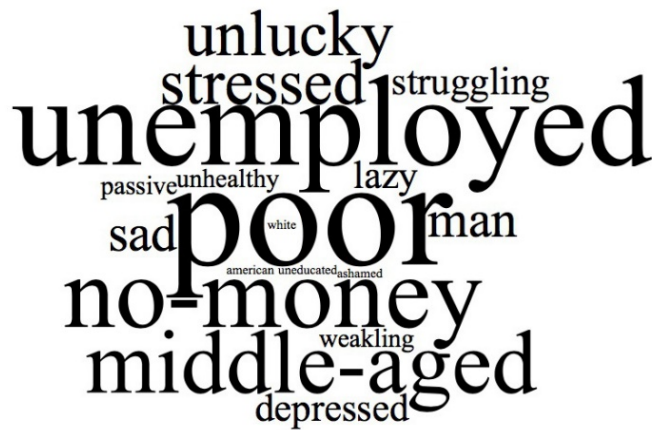


Figure 1. The descriptors of the poor version of the other player, from the free-text response in Study 1. The size of the word corresponds to its frequency, where larger size = more frequently existing word.

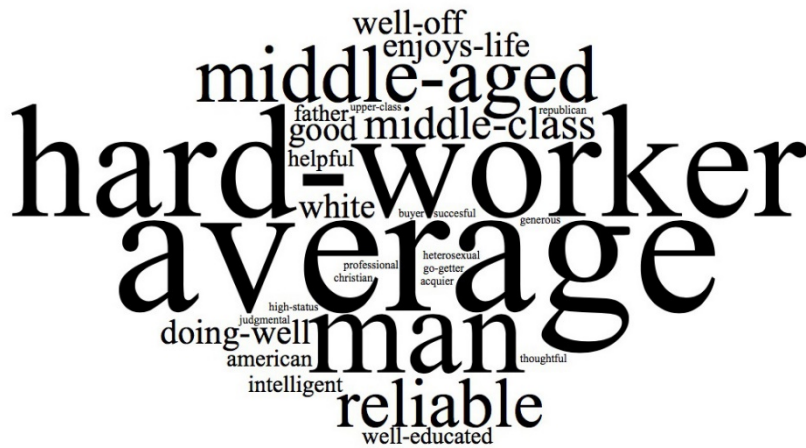


Figure 2. The descriptors of the nonpoor version of the other player, from the free-text response in Study 1. The size of the word corresponds to its frequency, where larger size = more frequently existing word.

Discussion

According to the SCM ratings of warmth and competence, we appear to have succeeded in activating the stereotype of the poor, and the notion of ‘the ordinary guy’ in the two different versions. The associations from the free-text responses confirm this. When comparing Figure 1 and Figure 2, it is obvious that the content of the associations clearly differed between the two versions.

When David was described as unemployed, with a SES of 2, the participants seemed to focus on him being unemployed and poor, with no money. In the version of him being full-time employed, with a SES of 6, the participants seemed to focus on him being average and hardworking. In both versions, the participants to some extent focused on David being a

middle-aged man. In the SES=6 version, they also thought of him as being reliable, while the SES=2 version activated descriptors such as him being unlucky and stressed.

Taken together, the results from the pilot study suggest that the information about the other player (either with low or medium-high SES) would function well as manipulation in the cooperation games.

Study 2: Trust game

In the main study, we wanted to analyze how the participants' decisions in the trust game differ depending on if they are playing with a poor vs. non-poor (i.e., having medium-high SES) person. If the tendencies described in the BIAS map (Cuddy et al., 2007) are correct, resulting in behavioral tendencies of passive harm towards the poor player, the predicted behaviors in relation to trust could manifest as tendencies to dismiss the situation of the poor version of the other player, with attitudes of pity or contempt. According to the BIAS map, 'the ordinary guy' (i.e., the non-poor version of the other player, corresponding to the stereotype of 'men'; HC/MW) should trigger passive facilitation, which for example can manifest as the tendency to include the person in team work and collaboration, resulting in attitudes of admiration (slightly higher warmth) or envy (slightly lesser warmth).

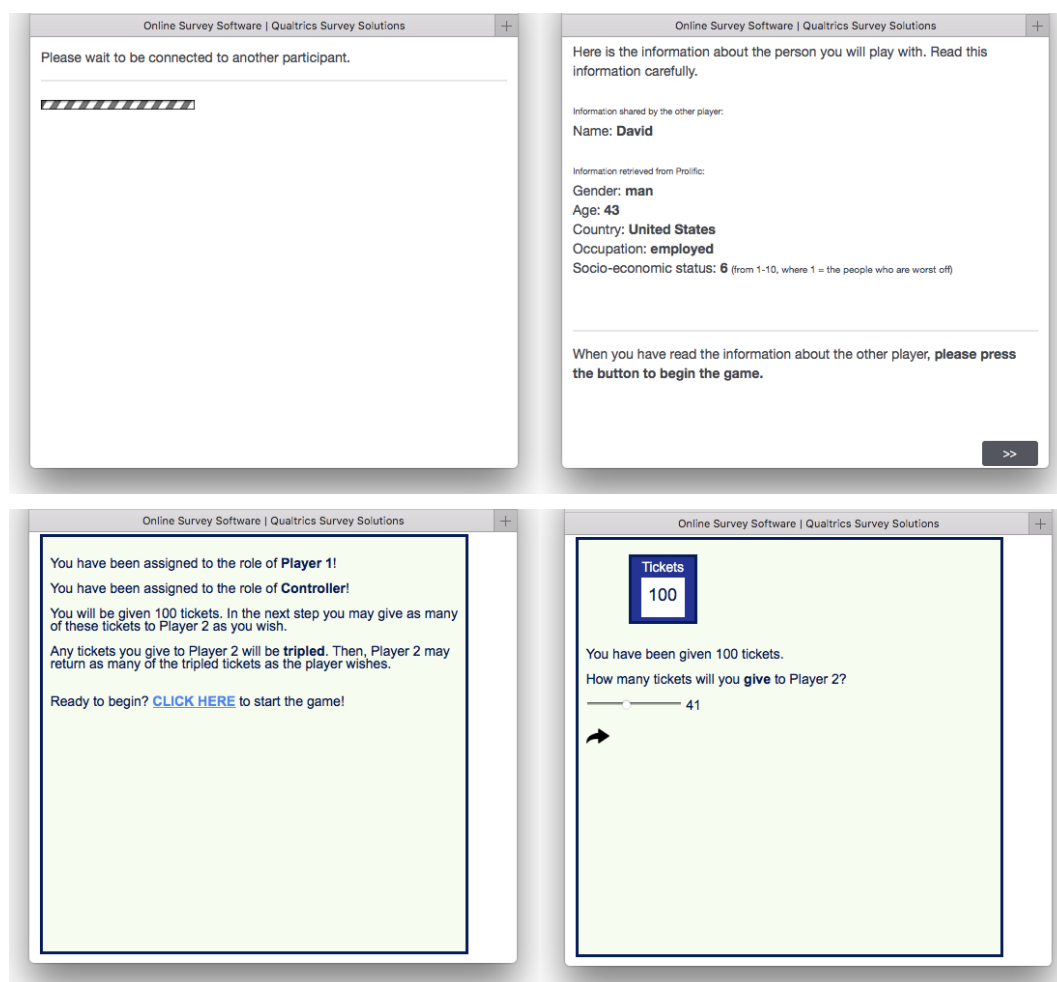
To increase participants' focus and motivation they played about real money (c.f., Rydval & Ortmann, 2004), and their reward for participating was determined by their score in the game. The trust game was performed online, using the survey platform *Qualtrics* (www.qualtrics.com), while the participants were recruited from the website *Prolific* (www.prolific.ac).

The other player. The trust game was embedded in a survey sequence, where participants were told they were going to play the game with another player online, while they actually played with the computer. After having read the rules of the game, the participants had to wait for eight seconds to be "paired" with another player. They were then asked to indicate their first name, and told that it would be presented to the other player, together with some demographics retrieved from Prolific. At the same time, they were going to see the corresponding information about the other player. The participants were then randomized to either read about the other player as having an SES of 2, or having an SES of 6. The information about the other player was the same as in Study 1, and can be found in Table 1. Figure 3 shows the interface of the game.

After having read one of the two descriptions of the other player, the participants were asked to indicate how they perceived him in terms of warmth (warm, sincere) and competence

(competent, confident). Ratings of the four items were made on a 5-point Likert scale where higher values imply higher warmth and higher competence, respectively (cf., the abbreviated questionnaire of SCM; Fiske et al., 2002). Then the game itself began.

*The game*². When the game started, the participants were told that they would be randomized to play the role as sender or as returner. They were also told that one of the players would be able to punish the other player during the game, this was also going to be decided by randomization. However, the participants always received the role as sender, and as punisher.



² The code for our trust game can be provided for anyone interested, just contact the corresponding author (anna.lindqvist@psy.lu.se).

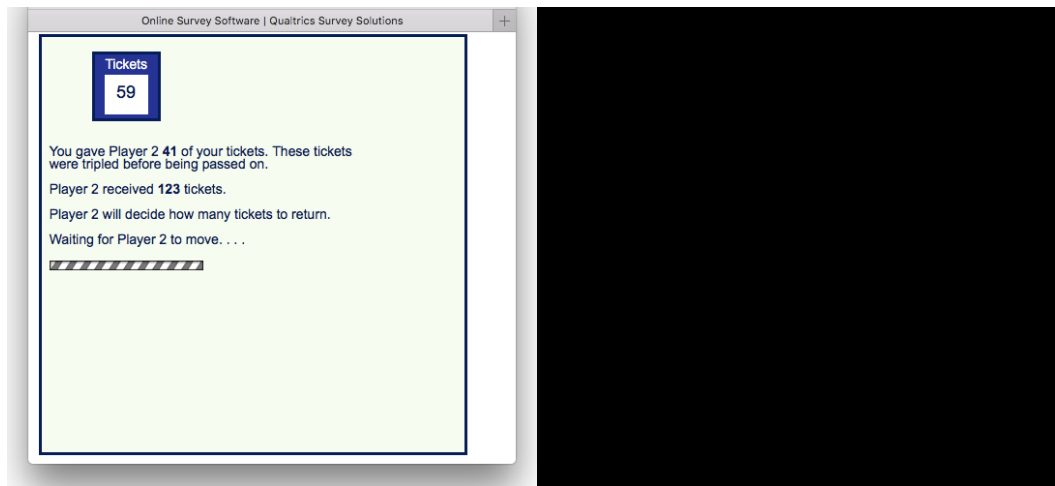


Figure 3a-f. The interface of the trust game: a) The participant is “waiting” for another player to connect; b) The stimulus/the information about the other player with medium-high SES; c) The information the participant gets when having been “randomized” to get the role as sender and as punisher; d) The interface where the participants chooses how many tickets to send to the other player; e) The participants is “waiting” for the other player to decide how many tickets he wants to return; f) The results from the game.

The participants/sender received 100 tickets, which they were told translates into \$10. Their first game task was to decide how many of these tickets they wanted to send to the returner/other player, i.e. to the computer, the sum was then being tripled. Next, the other player (i.e., the computer) returned a randomized number of tickets to the participant. The participants’ total score was hence the number of tickets they kept + the number of tickets the other player returned.

When the participants had received the returned amount of tickets, they had the opportunity to punish the other player by taking away 30% of his tickets. They were clearly informed that if they punished the other player, they would not get these tickets themselves. Both reward and punishment are seen as critical to regulating economic relationships, even in one-shot settings (Al-Ubaydli & Lee, 2012), where punishment can be interpreted as a negative form of reciprocity (Berg et al., 1995). By including the opportunity to punish the other player, we effectively added another (negative) measure of cooperation. The warmth dimension was hence measured in two ways: First by trust in terms of the amount of tickets sent to the other player, and secondly by the punishment.

Satisfaction. After finishing the game, the participants were asked to indicate their feelings about it. They did so by rating five items on a 5-point Likert scale, where higher values imply higher agreement with the statement: I am satisfied with my reward; The other

player collaborated well; The other player behaved egoistically; I was hoping to get a greater reward; I feel disappointed (the three last items are reversed).

Manipulation check. Lastly, the participants were asked to respond to two questions, used as manipulation check. First, we asked them to indicate what SES the other player had, on a scale from 1-10, to make sure they had paid attention to the information about the other player.

To make sure the participants had believed in our setup, i.e. that they were playing with a real person, not a computer, we also asked if the participants had recognized something about the other player that they wanted to express, as recommended by Belli and colleagues (2012). The participants could write anything they wished about the other player in a free-text response. This question also served as an opportunity for the participants to share their thoughts and feelings about their own and the other player's behavior in the game.

Participants

The 210 participants (self-identified as 90 women, 119 men, 1 did not indicate gender) between 18-69 years ($M = 35.07$, $SD = 11.84$) were recruited through the web platform *Prolific* (www.prolific.ac) where it is possible to retrieve relevant background demographics of the participants. We noted their country, personal income, and SES which is assessed on a 10-point scale from 1 = lowest possible SES to 10 = highest possible SES, with anchoring descriptions of the highest and lowest steps.

The participants were from the UK (94), US (77), Canada (10), New Zealand (4), China (3), Ireland (3), Croatia (2), Australia (2), Vietnam (2), Brazil (1), Colombia (1), Germany (1), India (1), Lebanon (1), Poland (1), Portugal (1), Puerto Rico (1), Russia (1), Singapore (1), Spain (2), Sweden (1), Taiwan (1), Venezuela (1). Their yearly median income was between £20 000–29 999, and their average self-rated SES was 5.01 ($SD = 1.71$).

Results

Manipulation check. Among the 210 participants, 175 (83.8 %) gave a response to the question if they had recognized anything about the other player that they wanted to express, where 80 of these simply responded “no”, or that they had not recognized anything. However, 8 of the participants (3.8%) stated that they had realized they were playing the game with a computer instead of a real person. These were excluded from the following analyzes.

The remaining 202 participants performed relatively well in remembering the other player's SES: $M_{poor} = 2.25$, $SD = 1.06$, $M_{nonpoor} = 6.02$, $SD = .59$. However, the average

remembered value of $M = 2.25$ differs significantly from the correct value of 2, $t(102) = 2.34$, $p = .021$. Four outliers, reporting SES values of ≥ 6 , were identified and excluded from the analyzes, resulting in a mean value of $M = 2.10$, $SD = .65$ which did not differ significantly from the correct value 2, $t(98) = 1.55$, $p = .123$. Among the participants who played with the non-poor version of the other player, the average remembered value of $M = 6.02$ did not differ significantly from the correct value of 6, $t(99) = .34$, $p = .733$.

Among the remaining 198 participants, 98 had been randomly assigned to play the game with the poor version of David, while 100 had been assigned to play with the nonpoor version. The SCM ratings of these versions corresponded to the earlier identified SCM content of the poor (cf. Lindqvist, Björklund, & Bäckström, 2017), and of the ‘average guy’ (i.e., ‘men’; A. J. C. Cuddy et al., 2009): There was no difference in warmth between versions, $M_{poor} = 3.34$, $SD_{poor} = .69$, $M_{nonpoor} = 3.31$, $SD_{nonpoor} = .68$, $t(196) = .325$, $p = .745$, $d = .04$. There was, however, a significant difference in competence ratings, $M_{poor} = 3.09$, $SD_{poor} = .67$, $M_{nonpoor} = 3.69$, $SD_{nonpoor} = .75$, $t(196) = 5.12$, $p < .001$, $d = .84$.

Also, when reading the free-text responses regarding whether the participants had recognized something about the other player, it became quite obvious that the manipulation had succeeded. Among those who played with the poor version of the other player, 48.0 % wrote something about him, where 51.0 % of these commented upon him being unemployed or having low SES. Among those playing with the nonpoor version, 40.2 % had something they wanted to share about the other player. Here, the responses were more varied compared to the poor version, mentioning his gender, commenting upon his strategy, upon his name, etc. Since no comment stood out particularly, we interpret this as a success in activating ‘an ordinary guy’ in the nonpoor version. All responses (except those just responding “no”, or the like, to the question) are shown in Appendix I.

Total score. On average, the participants’ total score in the game was 91.37 tickets ($SD = 44.47$) corresponding to a reward of \$ 9.4, ranging from 2–164 tickets. There was no significant difference in total score depending on if the participants played with the poor or the nonpoor version of the other player, $t(196) = .606$, $p = .545$.

Tickets sent. On average, the participants sent 57.28 ($SD = 30.44$) of their 100 tickets to the other player. This corresponds quite well to the amount noted by Berg and colleagues, where the participants sent \$5.16 of their \$10 (i.e., 51.6%). There was no significant difference in the number of tickets sent to the other player, depending on the other player’s SES, $M_{poor} = 54.71$, $SD = 30.02$, $M_{nonpoor} = 59.77$, $SD = 30.80$, $t(196) = 1.169$, $p = .244$, $d = .17$.

To analyze whether the number of tickets sent differed depending on the SES of the participants, we divided the participants into two groups of “low SES” (SES 1-3) and “medium-high SES” (SES 4-10). Thereafter we performed an ANOVA with number of sent tickets as dependent variable, and with SES of the other player (poor, nonpoor), and SES of the participant (low, medium-high) as independent variables. There was no main effect of participants’ SES on the number of tickets sent to the other player, $M_{low} = 59.60$, $SE_{low} = 5.07$, $M_{medium-high} = 55.93$, $SE_{medium-high} = 2.52$, $F(1,174) = .420$, $p = .518$, $\eta^2 = .002$, and no interaction effect of participants’ SES and the other player’s SES, $F(1,174) = .840$, $p = .361$, $\eta^2 = .002$.

To also analyze whether the number of sent tickets differed depending on the income level of the participants, we made a median split of the participants’ personal income, dividing them into the groups “low income” (< £29 999) and “high income” (> £30 000). Thereafter we performed an ANOVA with number of sent tickets as dependent variable, and with SES of the other player (poor, nonpoor) and participants’ personal income (low, high) as independent variables. There was no main effect of participants’ personal income on the number of tickets sent to the other player, $M_{low} = 55.75$, $SE_{low} = 2.63$, $M_{high} = 61.33$, $SE_{high} = 3.88$, $F(1,189) = 1.42$, $p = .235$, $\eta^2 = .007$, and no interaction effect of participants’ SES and the other player’s SES, $F(1,189) = .321$, $p = .572$, $\eta^2 = .002$. In other words, the number of tickets sent did not seem to be affected by the other player’s SES, or by the participants’ SES or income.

When analyzing the free-text responses (see Appendix I), we found a tendency where some of the participants playing with the poor version of the other player sometimes seemed to be more generous when sending tickets, because they felt pity for him. In the free-text responses about the other player, 28.6 % of the participants giving a comment about the poor version explicitly mentioned this, by somehow stating that “he was unemployed, so he needed the money more than I do” (see Appendix I for all responses). At the same time, 6.1 % of the participants giving a comment about the poor version of the other player expressed disapproval and aggression towards him, confirming that the stereotype of the ‘poor’ also might evoke and contempt. It shall, however, be noted that 12.0 % of the participants playing with the non-poor version of the other player also expressed disapproval or aggressions towards him; some of these comments might be interpreted as “envy” (as predicted by the BIAS map) such as “they are happy to not give back to somebody less well as them”.

Satisfaction. Not surprisingly, there was a significant correlation between satisfaction ($\alpha = .91$) and the participants’ total reward in the game, $r = .78$, $p < .001$. There was also a

significant correlation between satisfaction and the tickets the other player returned to the participant (in percent), $r = .84$, $p < .001$. There was, however, no significant difference in satisfaction between the participants playing with the poor or with the non-poor version of the other player, $M_{poor} = 3.51$, $SD_{poor} = 1.22$, $M_{nonpoor} = 3.23$, $SD_{nonpoor} = 1.34$, $t(196) = 1.547$, $p = .124$, $d = .22$.

To analyze the impact of participants' own SES on satisfaction, we performed an ANOVA with satisfaction as dependent variable, and with SES of the other player (poor, non-poor), and SES of the participants (low, medium-high) as independent variables. As expected, there was no main effect of the other player's SES on the participants' satisfaction, $F(1,174) = 1.152$, $p = .285$, $\eta^2 = .007$. There was, however, a main effect of participants' SES where those with low SES were significantly more satisfied with the game result, $M_{low} = 3.66$, $SE_{low} = .166$, $M_{medium-high} = 3.41$, $SE_{medium-high} = .167$, $F(1,174) = 4.82$, $p = .030$, $\eta^2 = .027$, but no interaction effect of participants' SES and the other player's SES, $F(1,174) = .003$, $p = .957$, $\eta^2 < .001$.

We also performed an ANOVA with satisfaction as dependent variable, and with income of the other player (low, medium-high), and income of the participant (low, high) as independent variables. There was no main effect of participants' income on their satisfaction, $M_{low} = 3.30$, $SE_{low} = .112$, $M_{high} = 3.51$, $SE_{high} = .165$, $F(1,189) = 1.11$, $p = .293$, $\eta^2 = .006$, and no interaction effect of participants' income and the other player's SES, $F(1,189) = .444$, $p = .506$, $\eta^2 = .002$.

Taken together, the satisfaction of the participants was dependent on their total score, the proportion that the other player returned to them, and their own SES.

Punishment. In total, 51 of the 198 participants (25.8%) decided to punish the other player. In the version with the poor player, 18.4% decided to punish the other player, while 33 % of the participants playing with the nonpoor version decided to punish him, the difference was significant, $\chi^2(1, N = 198) = 5.54$, $p = .019$. Among those playing with the poor version of the other player, there was no difference in punishment between participants with low or high income, $\chi^2(1, N = 94) = 1.83$, $p = .176$, or between participants with low or medium-high SES, $\chi^2(1, N = 84) = .022$, $p = .882$. The same was found among those playing with the non-poor version of the other player, with no difference in punishment between participants with low or high income, $\chi^2(1, N = 99) = .577$, $p = .447$, or between participants with low or medium-high SES, $\chi^2(1, N = 94) = .838$, $p = .360$.

There was a significant difference in satisfaction between the participants who decided to punish the other player, and those who decided not to, where those who punished the other

player also tended to be less satisfied with the game result, $M_{punishment} = 2.05$, $SD_{punishment} = .91$, $M_{nopunishment} = 3.82$, $SD_{nopunishment} = 1.07$, $t(196) = 10.63$, $p < .001$, $d = 1.72$. The difference is not surprising – if the participants were satisfied with their score, they tended not to punish the other player.

The punishment follows the tendency identified earlier, where the participants were more generous to the poor version of the other player while they showed more aggression towards the non-poor version, and to a significantly greater extent punished the nonpoor version of the other player compared to the poor version.

Discussion

The aim of this study was to test whether a game paradigm could be used to analyze bias-related real-life behavior in terms of discrimination against individuals representing the stereotype ‘poor’. Since stereotype content is suggested be a mix of the two dimensions of warmth and competence (cf., Fiske et al., 2002), where our focus was on behavior related to the warmth dimension, we used the *Trust game* (Berg et al., 1995), testing cooperation and trust. The stereotype in focus was ‘the poor’ (constituting of MW/LC; see e.g., Lindqvist et al., 2017), where the predicted behavior towards individuals representing the group ‘poor’, according to the BIAS map (Cuddy et al., 2007) is passive harm.

Our study confirms that computerized games simulating real players can be used for investigating real-life behavior: Among the 210 participants playing the trust game with a simulated player, 198 participants (94.3 %) were included in the analyses since they did not show any sign of realizing that they had played with a computer, and were sufficiently good in remembering relevant personal details of the other player (such as his SES, which was our manipulation).

Regarding the behavioral tendencies described in the BIAS map (Cuddy et al., 2007), we expected the poor version of the other player to be met by passive harm, while we expected the non-poor version to be met by passive facilitation. Notably however, passive behaviors are reflecting less obvious intentions, with the consequences of, for example, excluding members of a target group. Since neither the intentions nor the behavior is active, it might be difficult to identify such behavior in the game, at least as it is currently set up.

When turning to the free-text responses, there was a little sign of spontaneously reported discriminatory attitudes towards ‘poor’ individuals, since 28.6 % of the participants commenting on the poor version of the other player spontaneously reported that they wanted to be generous towards him, while another 6.1 % rather reported feelings of disappointment or

disapproval (i.e., as mentioned earlier, see Appendix I for all comments). When playing with the non-poor version, these tendencies were not reported. Instead, almost a third of the participants playing with the non-poor version of the player spontaneously reported feelings of aggression towards him in the free-text response, where some explicitly expressed feelings that can be interpreted as envy. Others indicated willingness to cooperate with the non-poor version of the player. Sometimes, the identified attitudes in the free-text reports have ingredients of enviousness.

When it comes to punishment, the participants showed a tendency to be relatively generous towards the poor version of the other player, since they to a significantly lesser extent punished the poor version compared to the non-poor version. This *might* be due to the pity some of the participants felt for the poor version, compared to the envy some of the participants felt for the non-poor version.

To sum up, since the suggested behavioral tendencies in relation to the warmth aspect tested in this experiment are passive (i.e., passive harm and passive facilitation), our results might point into the direction of confirming these predicted tendencies, but since the trust game tests active behavior, such interpretations should be made with care. To be able to fully confirm that the participants' behavior can be interpreted as 'passive harm' and 'passive facilitation', one needs to develop a real-life measurement (e.g., further developing a game paradigm) explicitly testing passive behaviors.

Concluding discussion and some advice for future research

We tested whether a computerized trust game simulating real players is useful as a tool investigating real-life behavior in relation to the warmth dimension of a stereotype. Our conclusion is that a game paradigm provides good opportunities for measuring real-life behavior in controlled social experimental settings. We believe that computerized collaboration games could be used to test the hypothesized behavioral tendencies of the BIAS map, in relation to the warmth aspect of stereotypes. To comprehensively test the map, one needs to develop a game explicitly testing passive behaviors, as well as include a game testing the competence dimension of each target stereotype.

For fellow research colleagues who want to adapt this approach, we have some advice on the road. We believe it is crucial to have a concrete setup check, which does not disclose the 'fake' player. Without such manipulation check, it is in fact impossible to claim that one analyzes real-life behavior (i.e., impossible to claim validity of the results). Our suggestion is to follow the routine of Belli and colleagues (2012) by including a free-text question at the

end of the experiment, where the participants are asked if they noticed anything particular about the other player that they wanted to express. After this, the participants should be properly debriefed.

It is of course also important to have a ‘fresh’ sample of participants, preferably without experience of similar game studies simulating real players. A participant who cancelled their participation should, for example, not be allowed to take the test again, this could be managed by controlling for unique IP addresses.

When testing how certain stereotypes affect decisions in the game, one also needs to include a manipulation check to confirm that the target stereotypes were activated. This can be accomplished by asking some questions about the other player. In the present research, we included a short SCM rating (to confirm the stereotype content as identified by Fiske et al., 2002), as well as specifically asking the participants about the other player’s SES.

There are, of course, other games testing collaboration in different ways. The trust game is an example of a trust dilemma (Van Lange, Joireman, Parks, & Van Dijk, 2013) where the highest outcome for the sender/the participant occur when all players cooperate. Other types of social dilemmas which can be tested in games include the chicken dilemma and the public goods dilemma.

The *Prisoner’s Dilemma* (Tucker, 1950) is probably one of the most well-known game paradigms. The game is a typical chicken dilemma (Van Lange et al., 2013): The two players are given the choice to either collaborate or defect (sometimes they also may choose to withdraw from the game). If both choose to collaborate, both get a medium sized reward. However, if one of them chooses to collaborate, while the other chooses to defect, the one choosing to collaborate gets nothing while the one choosing to defect gets the maximum reward.

In public goods dilemmas, the decision in the game can be interpreted as a choice between contributing or not to the public good (Van Lange et al., 2013). In this category, we find the *Give some* game (Dawes, 1980). In the original version of the game, developed by Dawes, each of five players can choose between keeping \$8 for themselves, or giving \$3 to each of the other players. These decisions are made simultaneously by all players, and if all of them give \$3 to each other, all end up with \$12 (4 x \$3). However, it is of course also possible for each participant to keep \$8 for himself.

Compared to other paradigms, the trust game which we use in this study is unique, since the maximum outcome for the sender/the participant comes about when both of them collaborate. When focusing on the behavior of the sender, it is also the only paradigm which

tests nothing but trust; games testing the chicken dilemma and the public goods dilemma also test how egoistically the players behave. To keep the design as clean as possible, we hence believe the trust game is among the most suitable paradigms for testing behavior in relation the warmth dimension of SCM (Fiske et al., 2002). However, as stated earlier, passive behaviors are difficult to capture. The ultimate paradigm for testing the BIAS map would hence include a game where, in addition to active behavior, ‘passive harm’ and ‘passive facilitation’ can be measured separately. In practical terms, this means that the game would provide data on both whether the participants neglect the welfare of someone (passive harm) and whether they explicitly choose someone for team work (passive facilitation). At the same time, the behavior should be a consequence where the participant acts for his own purposes, which simultaneously includes or excludes the other group (c.f., Cuddy et al., 2007). Until such a game is developed, it appears that the trust game is the most suitable for testing the BIAS map.

References

- Al-Ubaydli, O., & Lee, M. S. (2012). Do you reward and punish in the way you think others expect you to? *The Journal of Socio-Economics*, 41(3), 336-343.
- Anderhub, V., Engelmann, D., & Güth, W. (2002). An experimental study of the repeated trust game with incomplete information. *Journal of Economic Behavior & Organization*, 48(2), 197-216.
- Ashmore, R. D., & Del Boca, F. K. (1981). Conceptual approaches to stereotypes and stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 1-35). New York, NY: Psychology Press.
- Belli, S. R., Rogers, R. D., & Lau, J. Y. (2012). Adult and adolescent social reciprocity: experimental data from the trust game. *Journal of adolescence*, 35(5), 1341-1349.
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*, 10(1), 122-142.
- Bottom, W. P., Holloway, J., Miller, G. J., Mislin, A., & Whitford, A. (2006). Building a pathway to cooperation: Negotiation and social exchange between principal and agent. *Administrative Science Quarterly*, 51(1), 29-58.
- Brühlhart, M., & Usunier, J.-C. (2012). Does the trust game measure trust? *Economics Letters*, 115(1), 20-23.
- Cañadas, E., Rodríguez-Bailón, R., & Lupiáñez, J. (2015). The effect of social categorization on trust decisions in a trust game paradigm. *Frontiers in psychology*, 6.
- Cardenas, J. C., Chong, A., Ñopo, H., Horowitz, A. W., & Lederman, D. (2009). To what extent do latin americans trust, reciprocate, and cooperate? evidence from experiments in six latin american countries [with comments]. *Economia*, 9(2), 45-94.
- Carlin, R. E., & Love, G. J. (2013). The politics of interpersonal trust and reciprocity: an experimental approach. *Political Behavior*, 35(1), 43-63.

- Croson, R., & Buchan, N. (1999). Gender and culture: International experimental evidence from trust games. *The American Economic Review*, 89(2), 386-391.
- Cuddy, A. J., Fiske, S. T., & Glick, P. (2007). The BIAS map: behaviors from intergroup affect and stereotypes. *Journal of personality and social psychology*, 92(4), 631-648.
- Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J.-P., . . . Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, 48(1), 1-33.
- Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, 31(1), 169-193.
- Engle-Warnick, J., & Slonim, R. L. (2006). Inferring repeated-game strategies from actions: evidence from trust game experiments. *Economic theory*, 28(3), 603-632.
- Esses, V. M., & Dovidio, J. F. (2002). The role of emotions in determining willingness to engage in intergroup contact. *Personality and Social Psychology Bulletin*, 28(9), 1202-1214.
- Fiske. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology*, 4th ed., Vol. 2 (pp. 357–411). Boston: McGraw-Hill.
- Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in cognitive sciences*, 11(2), 77-83.
- Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of personality and social psychology*, 82(6), 878-902.
- Fiske, S. T., Xu, J., Cuddy, A. C., & Glick, P. (1999). (Dis)respecting versus (Dis)liking: Status and Interdependence Predict Ambivalent Stereotypes of Competence and Warmth. *Journal of Social Issues*, 55(3), 473-489.
- Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, 6(3), 347-369.
- Johnson, N. D., & Mislin, A. (2008). Cultures of kindness: A meta-analysis of trust game experiments. Available at Social Science Research Network: <http://ssrn.com/abstract, 1315325>.
- Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, 32(5), 865-889.
- Lindqvist, A., & Björklund, F. (2017). The stereotype of the poor and predictions of their economic behavior. *The Journal of Social Psychology*, 1-18.
- Lindqvist, A., Björklund, F., & Bäckström, M. (2016). *How Economic Behavior is Perceived Depending on the Socio-Economic Status of the Agent*. Paper presented at the 28th APS Annual Convention, Chicago, Illinois, USA.
- Reeder, G. D. (1993). Trait-behavior relations and dispositional inference. *Personality and Social Psychology Bulletin*, 19(5), 586-593.
- Rydval, O., & Ortmann, A. (2004). How financial incentives and cognitive abilities affect task performance in laboratory settings: An illustration. *Economics Letters*, 85(3), 315-320.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, 300(5626), 1755-1758.
- Sefton, M. (1992). Incentives in simple bargaining games. *Journal of Economic Psychology*, 13(2), 263-276.

- Smith, F. G., Debruine, L. M., Jones, B. C., Krupp, D. B., Welling, L. L., & Conway, C. A. (2009). Attractiveness qualifies the effect of observation on trusting behavior in an economic game. *Evolution and Human Behavior*, 30(6), 393-397.
- Tucker, A. W. (1950). *A Two-person dilemma*. Mimeo: Stanford University.
- Van Lange, P. A., Joireman, J., Parks, C. D., & Van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes*, 120(2), 125-141.
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17(7), 592-598.
- Wilson, R. K., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly*, 59(2), 189-202.
- Wojciszke, B. (2005). Affective Concomitants of Information on Morality and Competence. *European Psychologist*, 10(1), 60-70.
- Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the Dominance of Moral Categories in Impression Formation. *Personality and Social Psychology Bulletin* *Personality and Social Psychology Bulletin*, 24(12), 1251-1263.

Appendix I: Free-text responses to the manipulation check in the trust game

Did you recognize anything about the other player that you want to express?

Poor version

At socio-economic status 2, the other player needs the money more than me. The fact he kept more reflects this, and I don't feel bad about him doing so.

Being unemployed one must seize any opportunity (that harms no one). I have no problem with that.

David was unemployed, so he probably did not want to give away any of his tickets.

Goodness

He did not play optimally!!!

he is an unfair person

He is unemployed so i gave him most of the tickets.

He is very egoistic and dont knows the game, because i gave him all my reward to get half.

He played fairly

He probably needs money more than me

He was fair and cooperative

He was greedy, I gave him all my tickets and he gave me 10 haha.

He was too quick

He was unemployed

he was unemployed

He was unemployed and I wish he would have kept more for himself

He's been through hard times before, so he knows what it's like to be treated like scum

He's unemployed, just like me.

I do not wish to say anything to the other player

I felt that the other player needed the money more than me and if they didn't share I got no benefit from punishing them and may as well let them keep it as their need was greater.

I guess it's possible that my socio-economic status was shown as higher than his, which inclined him to behave unfairly?

I may have warm feelings about other people in general, but I have little trust in men.

I saw that they were unemployed and low economic status, so probably needed the money more than me.

I think they were probably in the same economic class as me, even though I work and they don't

If he isn't from a particularly wealthy background, he was certainly generous with the return

It was nice of them to return so we were even.

Just that they were unemployed so probably needed the money more than me

Male

More generous than was required

not as well off as me

Nothing in particular. Very generous.

Only, I suppose, that the other player was represented as being low status, presumably to see if that biased me towards mistrust.

opponents played very well

Seemed fairly generous

That they were old and were pretty poor

They are awesome

They are rather similar to me.

they gave me the same amount minus 1

They probably needed the money more than I do, I don't fault them for keeping the majority of tickets. It was my intention for them to get as many as possible because it does not harm me to lose money I do not own in the first place.

They seemed to have a similar socioeconomic status as me.

They seemed willing to collaborate.

they were unemployed and called David

They were unemployed and probably needed money, so they probably wouldn't have returned it.

They were unemployed which is why I gave them more tickets

They were unemployed. So I would never have wanted to punish them

Unemployed

Very fair, splitting his tickets

We had a few similarities such as over 40 and currently unemployed.

We have similar socio-economic status.

Non-poor version

alpha male

considering I gave up the initiative he treated me very poorly

David, 43 from US and employed

generous

Greed

Greed ruled his action. I am glad I didn't give him 100, which was my first thought.

Greed. Player 2 was unwilling to share.

He behaved poorly

He had the same name as my dad, and I thought that was familiar to me.

he is selfish person who only returned 12 tickets

He seemed quite fair

He split what I gave him completely evenly, which actually was in my favor a little... I couldn't ask for more.

He was an employed 40-something year old man with a job. I thought perhaps after he was given my profile details he would have been a bit more accommodating in the fairness of the tickets. But it's not all that surprising.

He was greedy

he was quite smart

He was unemployed.

He was very fair

He's greedy.

He's quite generous

his name was David

I felt he had a higher socioeconomic status

I recognized that his name is David, he is 43 years old, lives in the United States, and is currently employed.

lower socio status

Male and mature

Male, 43, 6/10 socio economic status and was employed (of memory serves)...shame

My confidence in middle-aged American males is not very high since the American election.

Near my age

No, nothing other than he was very fair.

Not really. I didn't really take the socio-economic status into consideration because I don't judge people based on that. I always give people the benefit of the doubt first if I don't know them and assume that they would too since both of us know the rules of the game.

Only sent 10 tickets....Greed much?

Same age as me

seemed like me but male

Selfish

Selfishness

similar to me

Their gender and social status were similar to mine.

Their name was David and they played quite well.

They appear to have taken the same approach as me i.e. we are both doing studies in order to earn some money and therefore played in order to get both of us the greatest reward

They are happy to not give back to somebody less well as them.

We were a similar age and from a similar socio-economic status