

Chapman University

Chapman University Digital Commons

ESI Working Papers

Economic Science Institute

12-2022

Predictive Mind Reading from First and Second Impressions: Better-than-chance Prediction of Cooperative Behavior

Eric Schniter

Chapman University, schniter@chapman.edu

Timothy W. Shields

Chapman University, shields@chapman.edu

Follow this and additional works at: https://digitalcommons.chapman.edu/esi_working_papers



Part of the [Econometrics Commons](#), [Economic Theory Commons](#), and the [Other Economics Commons](#)

Recommended Citation

Schniter, E., & Shields, T. W. (2022). Predictive mind reading from first and second impressions: Better-than-chance prediction of cooperative behavior. *ESI Working Paper 22-19*.
https://digitalcommons.chapman.edu/esi_working_papers/380/

This Article is brought to you for free and open access by the Economic Science Institute at Chapman University Digital Commons. It has been accepted for inclusion in ESI Working Papers by an authorized administrator of Chapman University Digital Commons. For more information, please contact laughtin@chapman.edu.

Predictive Mind Reading from First and Second Impressions: Better-than-chance Prediction of Cooperative Behavior

Comments

ESI Working Paper 22-19

**Predictive Mind Reading from First and Second Impressions:
Better-than-chance Prediction of Cooperative Behavior.**

Eric Schniter^{1,2,3,4*} and Timothy W. Shields^{1,3}

16 February 2023

Abstract: Could cooperation among strangers be facilitated by adaptations that use sparse information from first and second impressions to accurately predict cooperative behavior? If so, does more of this information lead to more accurate predictions? We hypothesize that predictions are influenced by stereotypes, descriptions, appearance, and contextualized behavioral history available for first and second impressions. We also hypothesize that predictions improve when more information is available. We conducted a two-part study. First, we recorded thin-slice videos of university students just before their choices in a repeated Prisoner's Dilemma (PD) with matched partners. Second, a worldwide sample of raters evaluated each player using either videos, photos, only gender labels, or neither images nor labels. Raters guessed players' first-round PD choices and then their second-round choices after reviewing first-round behavioral histories. Our design allows us to investigate the incremental effects of gender, appearance, and behavioral history gleaned during first and second impressions. Predictions become more accurate as treatments incrementally add gender, appearance, and behavioral history, but predictions from videos are no more accurate than from photos. These results demonstrate how predictive mind reading abilities function under sparse information conditions, helping explain why conditional cooperation is common among strangers.

Keywords: Mind reading, Cheater detection, Cooperation, Prisoner's dilemma, Photographs, Thin-slice video

Affiliations

¹ Economic Science Institute, Chapman University, Orange, CA 92866, USA.

² Center for the Study of Human Nature, California State University Fullerton, Fullerton, CA 92831, USA.

³ Argyros School of Business and Economics, Chapman University, Orange, CA 92866, USA.

⁴ Integrative Anthropological Sciences Unit, University of California-Santa Barbara, Santa Barbara, CA 93106, USA.

*Corresponding author email: eschniter@gmail.com

Manuscript word count without tables, figures, references: 7769 words

1. Introduction

Opportunities for cooperative interaction with strangers and for repeated interaction have presented recurrent adaptive problems over the course of human evolutionary history (Fehr & Henrich, 2003). Among our ancestors, potentially valuable interactions with strangers were often fraught with danger, exploitation, and mistrust (Daly & Wilson, 1988; Martin & Frayer, 2014, Wrangham, 2019). These consequences shaped our minds to detect and predict cheaters and cooperators in social contracts (Cosmides & Tooby, 1992; Green & Phillips, 2004). These adaptive problems continue to present themselves in modern society (Nowak & Sigmund, 2005; Seabright, 2010). Once reputations from interaction histories establish, partners can reap steady gains from iterated cooperation with one another (Andreoni & Miller, 1993; Kaplan et al., 2012, 2018; Kreps et al., 1982). However, established cooperators remain vulnerable to opportunistic exploitation by previously cooperative partners. Despite these challenges, humans will often cooperate with one another. In successfully navigating these mixed-motive social dilemmas, it has been proposed that people apply *mind reading* to intuit the beliefs and intentions of others and to predict their cooperation or cheating behavior (McCabe & Smith, 2001; Sylwester et al., 2012).

Far from a magical or telepathic ability, mind reading is an ability universal to all normally developing humans that people make regular use of (Whiten & Byrne, 1991). Mind reading is also known as mentalizing (Frith & Frith, 2006), applying a theory of mind (Premack & Woodruff, 1978), or using folk psychology (Dennett, 1989). When mind reading, people use available clues, evidence of past behavior, and context to update their beliefs and expectations about others, and to predict others' behavior in an instant (Baron-Cohen, 1997).

We test the general hypothesis that people use mind reading to rapidly predict behavioral propensities under sparse information conditions, such as upon first and second impressions of strangers, a fundamental social skill important for regulating cooperative and non-cooperative behaviors across novel and repeated

PREDICTIVE MIND READING

interactions. We also test the general hypothesis that predictions improve as more information is made available for first and second impressions. Below we explain our experimental approach and detail our predictions that mind readers inform their guesses about strangers by applying their prior beliefs and available clues revealed by the target's description, appearance, and behaviors.

We conducted a non-deceptive two-part study with financially motivated participants. In part one, across multiple rounds of play between matched partners, we recorded "thin-slice" videos only a few seconds in duration (Ambady & Rosenthal, 1993) showing face-and-shoulder closeups of a university sample of participants taken just before their choices in each round of a "Split or Take All" Prisoner's Dilemma (PD) game variant with unknown end-game. In the second part of our study, we recruited online a set of raters to first make guesses about the stereotypical male and female cooperation rates from the PD, then to guess the PD game behaviors of players. For each player guessed about, we provided a unique ID number and manipulated whether raters viewed either a thin-slice video showing the player's face, a photo still from the video, the player's self-identified gender label without photo or video, or only the ID (i.e., no gender label, no face from photo or video). After forming a first impression, raters guessed each player's behavior in the first round of gameplay. Raters also guessed behavior in the second round after viewing round one behavioral history and forming a second impression.

In social dilemmas like the repeated PD, pursuing short-term non-cooperative interests is at odds with the interests of developing cooperative partnerships. Despite the higher monetary rewards from successful non-cooperation, it has been demonstrated in social dilemma experiments that cooperation can develop with unrelated strangers in one-shot environments (Balliet & Van Lange, 2013; Dawes & Thaler, 1988; Dickhaut et al., 2008; Kiyonari et al., 2000; McCabe et al., 1996, 2003; Ostrom & Walker, 2003; Schneider & Shields, 2022), finitely repeated games (Andreoni & Miller, 1993; Dawes & Thaler, 1988; Embrey et al., 2018; Mao et al., 2017), and infinitely repeated games with unknown endgame (Camera and Casari, 2009; Duffy and Ochs, 2009, van den Assem et al., 2012; Normann & Wallace, 2012).

PREDICTIVE MIND READING

When a population of players contains a mix of cooperative and uncooperative types, mind readers might expect that those with a cooperative intention initially choose to cooperate and those with an exploitative intention initially choose to cheat. For conditional cooperators who prefer cooperating when their partner is a cooperator, beliefs about the ratio of cooperators to exploiter types in a population should be an important predictor of the strategies deployed in first-round interactions (Kiyonari et al., 2000). To predict players' cooperative propensity, mind readers must first infer or detect their intention. To do this upon first-impression, when no prior reputational information is available, mind-readers may apply their "homemade" prior beliefs about the ratio of cooperators to non-cooperators likely to be encountered (Camerer & Weigelt, 1998), or derived from stereotyped assumptions about targets (Ames et al., 2012; McCabe et al., 2000). For example, participants expect that most other people in experiments with them are cooperative (Andreoni & Miller, 1993; McCabe et al., 2000), consistent with cooperation rate evidence (Andreoni & Miller, 1993; Camerer & Weigelt, 1998; Hayashi et al., 1999; Kiyonari et al., 2000; Kurzban & Houser, 2005; McKelvey & Palfrey, 1992). From these considerations, we derive our first prediction: **(P1)** in the treatment where gender is not revealed, first-impression guesses of players' round 1 cooperativeness will be influenced by prior beliefs about generalized cooperation propensity in the player population.

People expect behavior in social dilemmas to vary by gender, and when players' gender is revealed, people expect gender to be predictive of strategic behavior (Fetchenhauer et al., 2010; Schniter & Shields, 2020; Sylwester et al., 2012). Across cultures, people stereotype others' tendencies to cooperate based on gender (Eagly, 2009). Prior research shows that males and females share gender stereotypes about cooperative, communal females and uncooperative, agentic males and that gender is one of the most accurate and universally agreed-upon stereotypes among commonly stereotyped social categories (e.g., age, gender, ethnicity, and social class) (Jussim et al., 2009; Löckenhoff et al., 2014). Gender is often inferred from appearance (e.g., in terms of culturally masculine and feminine attire, makeup, hairstyle), display of sexually dimorphic body features, behavior, and description (e.g., with symbols, labels, pronouns, and

PREDICTIVE MIND READING

proper names). Upon visual inspection, male and female gender is differentiated in less than a second (Fletcher-Watson et al., 2008), and usually achieving accuracy above 95% (Bruce et al., 1993; Bruce & Young, 2011; Hill et al., 1995; Jaeger et al., 2020). This suggests that descriptions and appearance revealing gender inform raters of gender-specific behavioral propensities that could be used for predicting PD strategies that males and females deploy in interactions with strangers. Of course, to successfully apply gender stereotypes to predictions of strangers' behavior, their gender needs to be known and stereotypes need to be accurate. We expect that **(P2)** in the gender-label treatment and treatments showing players' appearance—where gender is revealed, first-impression guesses of players' round 1 cooperativeness will be influenced by prior gender stereotyped beliefs about male and female players, and **(P3)** more accurate gender stereotypes will drive more correct guesses for treatments with gender labels and players' appearance than for the treatment with neither gender label nor appearance.

When faces can be seen in photos (Fetchenhauer et al., 2010; Tognetti et al., 2013), thin-slice video (Ambady et al., 2000; Ambady & Rosenthal, 1993; Fetchenhauer et al., 2010; Vogt et al., 2013), or during brief personal interaction (Brosig, 2002; DeSteno et al., 2012; Frank et al., 1993; Reed et al., 2012a), first impressions are formed using the static or dynamic clues encountered (Snyder, 1984). Faces may communicate information about stable dispositional traits like cooperativeness (Fetchenhauer et al., 2010; Frank, 1988; Frank et al., 1993), and distinguishing characteristics like gender, formidability, health, kinship, and ethnicity (Bruce et al., 1993; Fasolt et al., 2019; Zilioli et al., 2015). Facial displays of happiness and anger could also be helpful for predictive mind reading, as these displays are produced and understood by everyone, quickly interpreted—in well under a second (Batty & Taylor, 2003), and may be reliably informative of behavioral propensity (Ekman et al., 1987; Hirshleifer, 1987; Reed et al., 2012a; Verplaetse et al., 2007). As these clues can be diagnostic of cooperative propensity, and first impressions from appearances may sometimes be accurate (Fetchenhauer et al., 2010; Tognetti et al., 2013; Verplaetse et al., 2007; Vogt et al., 2013), when possible, mind readers are expected to use faces to improve guess

PREDICTIVE MIND READING

accuracy. Thus, we predict that **(P4)** first-impression guesses of round 1 cooperativeness will be more accurate in treatments showing a photo or video of players' faces than in treatments not showing faces.

Brief in-person interactions and thin-slice videos of only a few seconds may reveal dynamic information about players that static photographs cannot (Ambadar et al., 2005; Harwood et al., 1999; Pike et al., 1997; Sato et al., 2004). This dynamic appearance information may help mind readers make better predictions, but it could also present an unhelpful distraction. Dynamic faces may display “tells”, or involuntary facial cues, eye movements, blinking, and brief micro-expressions that can be used to assess the cooperative propensity of targets (Fetchenhauer et al., 2010; Frank, 1988; Frank et al., 1993; Hirshleifer, 1987; Reed et al., 2012). Dynamic faces may also reveal emotional expressivity, measured by the frequency and intensity of emotional expressions. Emotional expressivity can be used to index players' likelihood of cooperation, as more emotionally expressive faces tend to be more cooperative (Schug et al., 2010). While expressive behavior sampled in first impressions can improve judgmental accuracy (Ambady et al., 2000; Ambady & Rosenthal, 1993), it may not always be beneficial. Faces are highly arousing and provocative stimuli that may provide distraction that cannot be easily ignored (Palermo & Rhodes, 2007). If attention to faces is overly demanding of limited time or cognitive resources, the ability to make accurate behavior predictions upon first-impressions might be compromised. This possibility is consistent with distraction-conflict models of attention allocation (Baron, 1986; Durkin et al., 2020). Videos and in-person interactions that provide longer exposure to dynamic face stimuli may exacerbate this distraction problem. For example, Sylwester et al. (2012) asked raters to assess either thin-slice (1-5 seconds) or long (60-120 seconds) video clips of people playing a variation of the Prisoner's Dilemma (PD) game, and to predict whether each player would choose “Split” or “Take All”. Though they did not find that raters had above chance accuracy for long videos, they did find that accuracy was higher than expected by chance for thin slice videos. As the richer dynamic information from thin-slice videos may help form first and second impressions, we predict that **(P5)** guesses will be more accurate in the video treatment than in the photo treatment.

PREDICTIVE MIND READING

In repeated interactions prior demonstrations of partners' goodwill, such as costly helping or cooperative behavior, can help inform beliefs about their intentions to cooperate (Coricelli et al., 2000; McCabe & Smith, 2001). Even if first impressions are inaccurate, when new evidence of goodwill from cooperative behavior is revealed (e.g. after a round of game interaction), behavior predictions based on informed second impressions may become more accurate (Andreoni & Petrie, 2008; Schniter & Shields, 2014, 2020). Predictions of players' behavior may be made based on limited evidence of the goodwill players demonstrated in previous interactions. Players' willingness to pursue cooperation conditionally depends on preferences for mutual cooperation or exploitation, and consideration of whether partners previously demonstrated goodwill (Kiyonari et al., 2000). Conditional cooperators choose not to cooperate if their partners did not previously show goodwill, while non-cooperators are always expected not to cooperate. After round one, predictions of players' behavior should consider both players' and partners' previous show of goodwill. **Figure 1** outlines a goodwill accounting heuristic we hypothesize mind readers apply when predicting cooperative behavior. Mind readers should be able to evaluate the history of goodwill and then apply this simple heuristic quickly, with little cognitive effort or demand for additional information. Selection is expected to have strongly favored "fast and frugal" heuristics such as these because of their efficiency, inferential speed, and accuracy in decision-making situations constrained by limited information and available time (Gigerenzer & Goldstein, 1996; Todd, 2001). This leads us to predict that **(P6)** knowledge of past round behaviors in a repeated strategic interaction will improve the accuracy of guesses about players' future behavior.

Our paper proceeds as follows: in section 2 we review background literature and compare our predictive mind reading study design to others. In section 3 we provide methodological details, in section 4 we present results, and in section 5 we discuss the results, study limitations, and extensions.

PREDICTIVE MIND READING

2. Background

A cheater detection adaptation appears to have evolved for solving problems associated with social exchange and cooperation (Cosmides, 1989a; Cosmides & Tooby, 1989, 1992, 2005). Accurate detection and prediction of cooperators and defectors is crucial for avoiding the pitfalls of interacting with non-cooperators or missing opportunities with cooperators (Cosmides & Tooby, 2005; Frank, 1988). If different kinds of detection or prediction errors impacted fitness differentially, for example failing to identify a real threat vs. identifying a non-existent threat, then according to the “smoke detector principle” (Nesse, 2005) selection would have tolerated detection and prediction abilities with errors at different frequencies, favoring a bias towards more of the least costly type of error (Haselton et al., 2015; Haselton & Nettle, 2006). The design of our cheater and cooperator detection systems may reflect an error management bias toward overpredicting cheaters or, alternatively, towards overpredicting cooperators in ancestral populations (Barclay, 2008; Schaller, 2008; Sparks et al., 2016). However, it cannot be known whether selection favored one error bias over another without knowing more about ancestral base rates of cooperator and cheater encounters and the relative costs of detection errors for each type. Similarly, any contemporary evidence of detection bias and accuracy can only be evaluated in terms of modern rates of cooperation. Few studies offering support for accurate game behavior prediction from first impressions provide sufficient detail to compare cheater and cooperator detection abilities. Frank et al. (1993) found that after a conversation with a partner but before playing the PD, participants were able to accurately predict their partner’s behavior 69% of the time, with a greater improvement over chance seen for cheater detection than for cooperator detection. Brosig (2002) found that after a conversation with a partner but before playing the PD, participants were able to accurately predict their partner’s behavior 67% of the time, with a greater improvement over chance seen for cooperator detection than for cheater detection. Based on these results, there is no clear evidence for an error-prone bias or an enhanced ability to predict cheating relative to cooperation. In general, there is a mix of evidence both for and against abilities to predict cheating and cooperation; we review this below.

PREDICTIVE MIND READING

A fair amount of attention has been given to the evaluation of predictive mind reading abilities using the so-called “gold standard” (Funder, 2012) of better-than-chance accuracy, though it has been limited to predictions of targets with no reputational history of prior game behavior. Our study is unique in that we study not only round 1 guesses from first impressions, but also round 2 guesses of those same players, but from informed second impressions where raters know players’ behavioral history.

A few studies find support for accurate game behavior prediction (Brosig, 2002; Frank et al., 1993; Reed et al., 2012), however, others report mixed results with only partial support, or no support (Bonnefon et al., 2013, 2017; Efferson & Vogt, 2013; Fetchenhauer & Dunning, 2010; Jaeger et al., 2022; Kiyonari, 2010; Manson et al., 2013; Sparks et al., 2016; Sylwester et al., 2012; Tognetti et al., 2013; Verplaetse et al., 2007; Vogt et al., 2013). Some of these studies do not reward raters’ correct guesses (Sylwester et al., 2012; Tognetti et al., 2013; Verplaetse et al., 2007; Vogt et al., 2013), which may negatively affect the accuracy of raters’ guesses. In our study, correct stereotypes and guesses about individual players are incentivized with monetary rewards, which should motivate raters to reveal correct stereotypes and make their best guesses (Smith, 1976).

Many behavior prediction studies draw raters and targets from the same subject pool. In some cases raters were shown targets that they had prior interactions with or went on to play subsequent games with (Brosig, 2002; DeSteno et al., 2012; Frank et al., 1993; Manson et al., 2013; Reed et al., 2012a; Sparks et al., 2016). Our worldwide online sample of raters is not drawn from the same local communities as the players they guess about, neither from the same convenience samples as the players, nor from among the set of players themselves. While convenient, more insular designs invite the possibility that prediction results are confounded by raters’ prior familiarity with targets, their involvement in the subject pool or experiment session, or behavior norms specific to their local community.

PREDICTIVE MIND READING

Some have given attention to uncovering what aspects of targets' appearance might be helping people make behavior predictions (DeSteno et al., 2012; Jaeger et al., 2022; Manson et al., 2013; Reed et al., 2012a; Tognetti et al., 2013), though none of these have examined how well people can otherwise predict gameplay in the absence of personal cues from photos, videos, and face-to-face interactions, for example, by asking the question, "in the absence of visual stimulus, could strangers' gameplay be predicted with above-chance accuracy?". Our study design allows us to answer this question. Of the game behavior prediction studies that feature visual stimuli of players, many show images of the players under highly specific and unnatural conditions, such as where hair, clothes, and color are removed from faces or where faces are required to display emotionally neutral poses (Bonnefon et al., 2013; Jaeger et al., 2022). Other studies censor and manipulate the distributions of target characteristics to be equiprobable rather than varying naturally or representative of society's base rates (Oda et al., 2009; Olivola & Todorov, 2010). Yet other studies show videotapes of players, but drawn specifically from a setting unrelated to the game decision predicted (Brown et al., 2003; Fetchenhauer et al., 2010). Our study does not feature photos and videos from highly specific or unnatural conditions, nor does it censor or manipulate distributions of target characteristics. While our design controls the experimental settings and methods of stimulus capture, we allow PD participants to exhibit natural and ad libitum behavior in the contextually relevant moments before the PD game decision, when we capture their image.

3. Methods

3.1. Overview of experiments. Our study consists of two experimental procedures. In the first part, we use an experimental economic game and self-reported demographics to generate target stimuli consisting of thin-slice videos, facial photographs, identification numbers, gender labels, and behavioral strategies from a participant sample of game players. In the second part of our study, we use an economic experiment to ask whether raters can predict targets' game behaviors based on generalized cooperation beliefs, gender-specific cooperation stereotypes, static and dynamic appearance, and behavioral history.

PREDICTIVE MIND READING

3.2. *Stimuli (Prisoner Dilemmas)*. First, we conduct a computerized laboratory procedure in an experimental economics laboratory using a “Split or Take All” Prisoner’s Dilemma (PD) game variant with an unknown end-game and anonymous unacquainted matched pairs. We impose a random-stopping rule to terminate the game. We explained in the instructions that a computer algorithm determines the chance of players continuing to another round, and that players would interact a minimum of two rounds of game interactions with the possibility of more rounds.

Participants recruited to be ‘players’ in the PD were randomly drawn from a subject pool of graduates and undergraduates at Chapman University. We used no deception and paid these players for the outcomes of their behavior in the study. As such, all game decisions were incentivized by the economic consequences of the game. We ran 13 sessions, each taking approximately 60 minutes.

In this PD each player chooses between “Split” or “Take All” strategies. Players were provided payoff matrixes explaining the consequences of both players’ choices. Payoffs in the game are set up such that the reward from choosing “Take All” while one’s partner chooses “Split” is the largest payoff possible –the entire jackpot, greater than the payoff for a mutual “Split” –half of the jackpot. The mutual “Split” payoff is greater than the mutual “Take All” payoff of zero, which is equal to the payoff of zero from choosing “Split” while one’s partner chooses “Take All”. The strategy labels used are intuitive because they directly describe the payoff goals.

96 players aged 18 to 25 (51 men, 45 women) gave permission to be video recorded at intervals throughout the experimental procedure under standardized videographic conditions and for their recordings and experiment data to be made available for later research. Players were told that at no time would their or other players’ identities or video recordings be revealed to participants in their experiment session.

PREDICTIVE MIND READING

Videos of players were taken using computer display-mounted digital cameras in individual computer terminal cubicles, set at the same distance from uniform backgrounds. From the original video recordings capturing head-and-shoulder closeups with ad libitum behaviors and expressions in the moments directly preceding a game decision being made, we trimmed thin-slice videos two to three seconds in length without audio. Photographs showing each player's face were captured from the thin-slice video. Additional details of our PD game, procedure, and stimulus development are available online (<https://doi.org/10.5281/zenodo.4321821>).

3.3.1. Prisoner Dilemma Prediction Experiment. Between April 2021 and August 2022, we recruited 445 participants ($M_{\text{age}}=33.6$, $SD_{\text{age}}=12.0$; 48.53% male, 48.98% female) using www.prolific.co. Participants were allowed up to 87 minutes to complete the experiment. We restricted recruitment to volunteers residing in the US, and only allowed volunteers to participate in the study once. 422 participants remained after excluding participants for violating requirements; specifically, we excluded (i) 11 for taking the survey on a smart phone despite prohibition against using small screen devices, and (ii) 12 for completing the task in less than 480 seconds, a speed we considered to be humanly improbable. **Table 1** reports the characteristics of these participants whom we refer to as raters.

All raters received instructions. To advance to the prediction study, raters had to complete, without error, a series of control questions verifying that a human responder is attentive to our questions. Instructions and survey questions are available in the online Appendix B.

Raters received the same instructions for the Prisoner's Dilemma game that were provided to players in the first experimental procedure. Raters were informed that they would first make guesses about the round 1 behaviors of the set of female players and the set of male players in the original study. For example, "On a scale ranging from 0% to 100% of the time, how often do you guess that females chose to "Split" and "Take All" in the first round of the original experiment", with the requirement that these percentages must equal

PREDICTIVE MIND READING

100%. Raters answered identical questions about males. These guesses inform us of raters' female stereotype beliefs and male stereotype beliefs. Next, raters made a series of guesses about the game behaviors of each player from the original study by selecting either the cooperative strategy ("Split") or the uncooperative strategy ("Take All") that they expected the player to have chosen. Each rater made these guesses about each of the 94 players. First, all guesses about round 1 game behavior were made. Next, with the history of each player and partner's round 1 behavior provided, raters made all guesses about round 2 game behavior. We used no deception and paid raters for the accuracy of their guesses in the study. As such, all guesses made were incentivized by the economic consequences of their accuracy.

3.3.2. Treatment groups. We conducted a 4 x 1 between-subjects design with raters. Raters were randomly assigned to one of the treatment cells. There are four treatments manipulating player information available for first and second impressions. We call these "None" (n=108), "Label" (n=101), "Photo" (n=108), and "Video" (n=105). All treatments make player IDs available and manipulate behavioral history information within-subjects. No history is available for raters' first-impression round 1 guesses and the history of players' and partners' round 1 choices is available for raters' second-impression round 2 guesses. The Label, Photo, and Video treatments make gender information available. The Photo and Video treatments make static facial and appearance information available. Only the Video treatment makes dynamic face and appearance information available. As such, this design incrementally manipulates availability of information about gender, faces, static appearance, dynamic appearance, and contextualized behavioral information (see **Figure 2**), allowing us to systematically evaluate the general hypothesis that availability of more of this information for first and second impressions leads to better predictions. This design also allows us to test predictions about the role of prior beliefs (**P1**) the role of player gender (**P2**, **P3**), face appearance (**P4**), static vs. dynamic appearance (**P5**), and behavioral history (**P6**). We preregistered our treatments at aspredicted.org (#61202, #103594) before collecting their data.

PREDICTIVE MIND READING

3.4. *Consent and data availability.* Internal review board approval was granted by Chapman University (#1718H016, #1314H065). The data for statistical analyses are available at <https://doi.org/10.5281/zenodo.7465288>.

3.5. *Measurements and analyses.* To evaluate raters' abilities to predict player behaviors, we measure the correctness, cooperator and cheater detection rates, accuracy, bias, and odds-ratios of raters' guesses. We define 'correctness' as the proportion of observations for a rater where cooperators (those choosing "Split") and cheaters (those choosing "Take All") are correctly guessed. Likewise, 'cooperator detection' measures the proportion of observations for a rater where "Split" was guessed given the player chooses "Split", and 'cheater detection' measures the proportion of observations for a rater where "Take All" was guessed given the player chooses Take All. We define 'accuracy' as the function of a rater's cooperator detection rate (H) and cheater detection rate (R). Formally, this is defined as $[Z(H) - Z(1 - R)]$. The function $Z(\cdot)$ is the inverse of the standard normal cumulative distribution, where rates are greater than zero but less than one (Macmillan & Creelman, 2004). The accuracy function converts probabilities into z scores. Rates of zero (one) are transformed to $1/100$ ($99/100$) so that the z scores are more (less) than negative (positive) infinity. Bias is a measure of the rater's tendency to over- or under-predict actual cooperator or cheater rates. Formally, this is defined as $-0.5 [Z(H) + Z(1 - R)]$. Negative values represent a bias toward predicting cooperation, and positive values represent a bias toward predicting cheaters. The 'odds-ratio' of guesses is the number of correct guesses in a round divided by the expected number of correct guesses due to chance alone, where α is the proportion of time that the rater guesses "Split", β is the proportion of the time that players choose "Split", and the expected number of correct guesses due to chance is $\alpha\beta + (1 - \alpha)(1 - \beta)$. Better-than-chance accuracy is defined by odds-ratio values greater than one.

We use logit analysis to evaluate the effects of stereotypes, labels, static appearance, and dynamic appearance on raters' first- and second-impression predictions, correctness, cooperator and cheater detection rates, accuracy, and bias. For round 1 guesses about unknown gender players in the mixed gender

PREDICTIVE MIND READING

population, we calculate raters' 'generalized cooperation beliefs' from an average of their male and female stereotyped beliefs. To assess whether raters with more accurate gender stereotypes make more pre-impression correct choices, we create dummy variables for gender stereotype accuracy that we call the 'sufficiently correct median'. The dummy is one if the rater's gender stereotype is greater than or equal to 50% and players of that gender are cooperative, or if the rater's gender stereotype is less than 50% and players of that gender are non-cooperative. Otherwise, the dummy is zero. If raters base guesses only on correct stereotypes, then there should be an increase in correctness but not accuracy.

All models presented in this study can be replicated using the statistical and data files presented as electronic supplementary materials. Statistical analysis was performed using Stata/SE 17.0.

4. Results

Among PD players we can observe the endogenous emergence and natural distribution of cooperative behaviors among matched pairs and the effects of game interaction outcomes on subsequent game behavior.¹ Below we describe the results of our PD prediction study which elicited raters' gender stereotyped beliefs about male and female cooperativeness, followed by a series of predictions about individual PD players' game behavior based on first and second impressions. On average, raters completed the study procedure in 24.4 minutes and earned \$4.56. Prediction response times per target by treatment are reported with rater demographics in **Table** .

4.1. Stereotyped beliefs about PD players' cooperation rates

¹ Game data is available at <https://doi.org/10.5281/zenodo.4321821>.

PREDICTIVE MIND READING

Raters' generalized cooperation beliefs indicate that they expect players will cooperate 54 percent of the time in the first round (**Table 2**).² Males were stereotyped as less cooperative (44.2%) than females (63.9%). Gender-specific stereotypes about male and female players were heterogeneous (**Figure 3**), significantly correlated (Pearson 0.503, $p < .001$), and significantly different (Wilcoxon matched-pairs signed-rank test, $Z = 16.1$, $p < .001$). These stereotype beliefs significantly underestimated actual male player cooperation (61.4%) and female player cooperation (86.0%) in the first round (Wilcoxon, males: $Z = -14.3$, $p < .001$; females: $Z = -17.5$, $p < .001$). Male raters stereotyped males to be slightly more cooperative (46.3%) than female raters did (42.2%), a significant difference (Wilcoxon, $Z = 2.29$, $p = .022$). Similarly, female raters stereotyped females to be more cooperative (66.5%) than male raters did (61.8%), a significant difference (Wilcoxon, $z = 2.00$, $p = .045$).

² There were no significant differences in the generalized cooperation belief between treatments (Kruskal-Wallis, $p = .114$). Male and female generalized cooperation beliefs did not differ significantly (Kruskal-Wallis, $p = .300$)

PREDICTIVE MIND READING

4.2. First-impression guesses about PD players' round 1 game behavior

Upon exposure to stimulus describing and sometimes showing PD players deciding how to play in round 1 of a repeated PD, raters made rapid first impressions and predictions of each of 94 players, averaging across conditions 1.2 (None), 1.4 (Label), 3.1 (Photo), and 8.3 (Video) seconds per player. Consistent with their stereotypes, raters underestimated round 1 cooperation in all treatments, predicting 58.7 percent cooperation, when it was 74.5 percent (**Table 3, panel A**). Despite underestimated cooperation, raters show significantly better cooperation detection than cheater detection (Wilcoxon matched-pairs signed-rank test, all $p < .041$). Round 1 correctness, cheater detection, and cooperator detection rates are reported in **Table 3**. There were significant differences in guesses between treatments (Kruskal-Wallis, $p < .001$). Correctness and detection rates showed significant differences between treatments (Kruskal-Wallis, all $p < .001$), as does accuracy (Kruskal-Wallis, $p < .001$), and bias (Kruskal-Wallis, $p < .001$). Round 1 guess correctness shows no difference between conditions that do or do not reveal gender (Wilcoxon, $p = .259$). In gender conditions, cheater detection rates are significantly higher ($p < .001$), while cooperator detection rates are significantly lower ($p = .005$). In conditions revealing players' faces, round 1 cheater detection rates are significantly higher (Wilcoxon, $p < .001$), while guess correctness and cooperator detection rates are significantly lower (Wilcoxon, both $p < .001$). The odds-ratios in the Label, Photo, and Video treatments indicate better-than-chance accuracy (**Figure 4**) and are significantly higher than in the None treatment (Wilcoxon signed-rank test, all $p < .003$). In all treatments, the average bias is towards overpredicting cheaters (Wilcoxon, all $p < .043$).

Below, we evaluate our research questions concerning the predicted effect of stereotypes, labels, photos, and videos on guesses of round 1 game behavior.

Where gender cannot be detected, are round 1 guesses influenced by prior beliefs about generalized cooperation propensity in the player population (P1)? Yes.

PREDICTIVE MIND READING

In the None treatment, where the rater could not determine a player's gender, raters generally *believed* that players would cooperate 54.0% of the time and *guessed* that 58.7% of players would cooperate in round 1. The effect of generalized beliefs on guesses is significant in the None treatment ($p < .001$) (Table 4, regression 1).

Are round 1 guesses influenced by gender stereotypes in treatments where player gender is labeled or seen (P2)? Yes.

The effect of gender-specific stereotypes on guesses is positively significant in Label, Photo, and Video treatments where gender can be visually detected (all $p < .001$). (Table 4, regression 2). The effect of gender-specific stereotypes is significantly stronger for the Label treatment than the Photo treatment ($p < .001$), and the effect is significantly stronger for the Photo treatment than for the Video treatment ($\chi^2(1) = 258.98, p < .001$).

Is the correctness of round 1 guesses influenced by the accuracy of male and female stereotype beliefs in treatments where gender is labeled or seen (P3)? Yes.

Using a logit analysis, we find that more accurate stereotypes are correlated with more correct guesses ($\chi^2(7) = 483.18, p < .001$; BIC = 53,302). We interact treatment and sufficiently correct stereotypes on guess correctness for each of the 422 raters 94 first round guesses. For all treatments, sufficiently correct stereotypes are positively significantly correlated with correct guesses (all $p < .001$). The effect of sufficiently correct stereotypes on correct guesses is significantly stronger for Label treatment than for the Photo treatment ($\chi^2(1) = 14.78, p < .001$) or the Video treatment ($\chi^2(1) = 29.83, p < .001$). The effect of sufficiently correct stereotypes on correct guesses is marginally stronger for the Photo treatment than for the Video treatment ($\chi^2(1) = 3.04, p = .081$). The effect of sufficiently correct stereotypes on correct guesses is not statistically different in the Video treatment than the None treatment.

PREDICTIVE MIND READING

Are round 1 guesses more accurate in the treatments showing the player's face (P4)? Yes.

Guesses in the Photo and Video treatments showing the player's face are significantly more accurate than in treatments not showing the face (Wilcoxon, $p = .013$). Bias was significantly closer to neutral (zero) in treatments showing the player's face (Wilcoxon, $p < .001$).

PREDICTIVE MIND READING

Are round 1 guesses more accurate in the video treatment than the photo treatment (P5)? No.

Accuracy in the Video treatment is significantly lower than in the Photo treatment (Wilcoxon, $p = .045$) and the Label treatment (Wilcoxon, $p = .019$). Bias was not significantly different between treatments showing the player's face.

4.3.1. Second-impression guesses of PD players' round 2 game behavior

Raters guessed 58.8% of players would cooperate in round 2, quite close to their guess of 58.7% cooperation in round 1 (**Table 3, panel A**). Players' cooperative behavior decreased from 74.5% in round 1 to 53.2% in round 2. Compared to round 1 guesses, correctness improved 5.3%, cheater detection improved 7.8%, and cooperator detection improved 9.1% for round 2 guesses (**Table 3 panel A**). Below we report results that help explain these performance improvements.

Does knowledge of past round behavioral history improve the accuracy of guesses (P6)? Yes.

Across treatments, round 2 guesses were more accurate than chance (**Figure 3**) (Wilcoxon, all $p < .001$). The accuracy of the round 2 guesses (0.595) increased above the accuracy of round 1 guesses (0.124) (**Table 3, panel A**). In every treatment, the increase in accuracy was significant (Wilcoxon signed-rank test, all $p < .001$). Next, we conduct post-hoc analysis to determine whether the artifactual conditions endogenously created by players' round 1 behavioral history affected raters' round 2 guess performance, and how gender stereotypes or players' facial appearance may have played a role.

4.3.2. Post-hoc analyses of round 2 guesses given stereotypes, conditions with gender or faces revealed, and behavioral history.

PREDICTIVE MIND READING

Though stereotypes are closer to round 2 than to round 1 player behavior, across behavioral histories, stereotypes and knowing the player's gender affect round 2 guesses significantly less than round 1 guesses (**Table A1**). Round 2 guess correctness, cheater detection, and cooperator detection show no relative improvement in gender conditions (Wilcoxon, $p = .140$, $p = .091$ and $p = .079$, respectively). However, in conditions revealing players' faces, round 2 guess cheater detection rates are significantly higher (Wilcoxon, $p = .010$), while cooperator detection rates are significantly lower (Wilcoxon, $p = .014$) and there is no significant difference in correctness (Wilcoxon, $p = .080$).

Raters' round 2 guesses and correctness vary across the players' four possible behavioral histories: 'Both Take All', 'Take All/Partner Split', 'Split/Partner Take All', and 'Both Split' (**Table 3, panel B**). Compared to the round 1 guesses, raters significantly increased their round 2 guesses of cooperation for the 'Both Split' behavioral history condition, and significantly decreased their guesses of cooperation for behavioral history conditions where at least one partnered player chose "Take All".

Round 2 guesses are affected by seeing gender labels or faces in the context of round 1 behavioral history, as these cues help improve guess correctness about male players generally (**Table A1**), and guess correctness for all players in the behavioral conditions where one or both partners chose "Take All" (**Table A2**). We find significant differences in correctness when raters had access to player's faces, but no significant differences between the Photo and Video treatments. When raters see a player's face in a 'Take All/Partner Split' interaction, they more aptly detect whether the player will be a round 2 cheater. Likewise, when raters see the player's face in a 'Split/Partner Take All' interaction, they more aptly predict that the player will choose 'Split'—resulting in more correctness than in conditions without the player's face visible (Table A2).

5. Discussion

PREDICTIVE MIND READING

These results provide supporting evidence for mind reading mechanisms designed to rapidly predict others' cooperativeness when forming first and second impressions. Below we discuss the importance of beliefs, personal and contextual clues, and methodological approaches for revealing predictive mind reading abilities.

Our results suggest that mind-reading does not wait for strategic interaction to begin. When the incentive structure of a game is easily understood and in the absence of direct behavioral evidence, raters can make these predictions easily and in rapid succession, taking about 3 to 4 seconds on average to evaluate and guess about each player. The incentive structure we chose for our "Split or Take All" PD game is one that we expect can be widely understood – leading to common perspectives and expectations among players and raters; it is identical to that of games featured on television shows such as *Friend or Foe*, *Golden Balls*, or *Take It All* which have since been analyzed as a natural experiment of cooperation (Burton-Chellew & West, 2012; van den Assem et al., 2012). In *Friend or Foe*, *Golden Balls*, or *Take It All* games, players choose "Split" 53 percent of the time, and young adult males are less cooperative than young adult females, though the gender effect reverses for older players (Burton-Chellew & West, 2012; van den Assem et al., 2012). Our raters expected males to be less cooperative, and for players to cooperate 54 percent of the time, almost identical to the game show average.

Raters in our study appear to rely on gender stereotyping to predict strangers' future behavior. With sufficiently correct gender stereotype beliefs and knowledge of the players' gender or appearance, raters make better-than-chance guesses of unacquainted players' round 1 cooperation behaviors in a repeated PD. Interestingly, the effect of correct gender stereotypes on correct guesses of a player's round 1 behavior is strongest in the gender label treatment.

Most studies evaluating the role of gender in raters' evaluations of target's cooperativeness rely on raters inferring gender from visual inspection and provide no control for whether something about appearance

PREDICTIVE MIND READING

other than gender might be confounding the effect attributed to target gender (Fetchenhauer et al., 2010; Sylwester et al., 2012; Tognetti et al., 2013; Vogt et al., 2013). We created a gender-label treatment to carefully isolate the effect of male or female gender from other effects of visual appearance. While all our PD players self-identified as either male or female, a small portion of our raters chose to not identify as male or female. Future studies will benefit from inquiry into the alternative gender identities and concepts that are becoming increasingly preferred by survey respondents and might better reveal gender influences if carefully measured (Snyder et al., 2022).

The informational differences afforded by our treatments suggest that raters may not have equal reason to rely on gender stereotypes across treatments. Across conditions where gender can be visually detected, the Label treatment provides raters less player information than the Photo treatment, which provides less player information than the Video treatment. As a result of these differences in available information, raters may trade off the value of gender clues for additional visual clues. An additional concern about differences across these treatments is that the appearance of static or dynamic faces may present an unhelpful distraction for raters who might be better off relying on accurate prior beliefs. The formation of first impressions from faces may be so automatic and non-conscious that they are relied upon even when objectively better information is available (Olivola & Todorov, 2010; Rezlescu et al., 2012) or when it is known that one should avoid being influenced by faces (Blair et al., 2004; Hassin & Trope, 2000; Palermo & Rhodes, 2007). While raters in our study appear to be trading off the influence of correct gender stereotypes for additional appearance information, the effect of more appearance information on guess correctness and accuracy is negative: the Photo treatment is less correct and accurate than the Label treatment, and the Video treatment is even worse off, consistent with the conflict-distraction model.

Predictions in the Video treatment are no better than predictions in the Photo treatment, indicating an exception to the general result that prediction performance benefits from more available information for first and second impressions. Visual and auditory attention to dynamic faces (e.g., changing eye gaze,

PREDICTIVE MIND READING

expressions, utterances) requires more time and attentional resources potentially distracting or interfering with processing capacity for tasks separate from visual and aural inspection (Lavie, 1995; Pessoa et al., 2002). The attentional costs and longer response times (about 5 additional seconds per target) over Photo in our Video treatment may have contributed to a greater conflict-distraction effect, producing less guess correctness and accuracy than in the Photo treatment. Though our research design did not compel standard response times across treatments to control for these costs, future research may benefit from designs with response time controls or from the investigation of response time variation.

Upon learning the details of players' round 1 PD interactions, raters can form updated second-impressions with the new information gleaned. From these second impressions, raters make better-than-chance predictions of players' round 2 PD game behaviors across all treatments – improving their guess performance from round 1 guesses. Additionally, the treatments showing players' faces improve cheater detection but hurt cooperator detection. Correct and accurate round 2 guesses are also affected by behavioral history conditions where one or both partners chose “Take All”. Prior research suggests that more masculine male faces are associated with perceptions of aggressiveness and dominance (Geniole et al., 2015; Sell et al., 2009; Zilioli et al., 2015), consistent with the idea that males who appear stronger and more masculine have greater potential bargaining power via coercive formidability and therefore can be expected to act more aggressively, reactively, and less cooperatively in social dilemma interactions (Daly & Wilson, 1988; Sell et al., 2012). Given our effects of male faces on round 2 guesses, it may be productive for future research to investigate further how variation in male cues, such as facial masculinity and formidability, may be predictive of cooperativeness in repeated games, especially in the context of previously non-cooperative interactions where entitlement and reactive anger may be at play.

Combined, cheater and cooperator detection abilities provide raters better-than-chance behavior prediction abilities in most of the treatments for round 1 guesses, and in all the treatments for round 2 guesses. When cooperator and cheater detection are compared, we see that cooperators were more accurately detected than

PREDICTIVE MIND READING

cheaters. Across round 1 and round 2 guesses, cooperator detection is higher in the gender label condition than in the conditions showing players' faces, and cheater detection improves in conditions showing players' faces. The bias to over-predict cooperation in our study might reflect the error management associated with social dilemma problems in hyper-cooperative modern society: where the costs of not detecting cheaters and exposing oneself to interactions with them are small relative to the gains of mutual cooperation with cooperative partners (Delton et al., 2011; Sparks et al., 2016). If our adaptations for social exchange have been sensitive to changing costs of being cheated versus missing cooperation opportunities, the calibration of our error management for cheater versus cooperation detection might be facultative rather than fixed – responding to serve the more costly challenges in the local social exchange environment (Barclay, 2008). Our results also suggest that cheater and cooperator detection abilities respond to sparse cues, like gender and appearance, available in first and second impressions.

As gender identity and photo or video appearance are influential parameters in self-presentation across a variety of human interaction mediums affecting investment, voting, legal decisions, hiring, mate selection, and cooperative interaction (Snyder et al., 2022; Todorov, 2017), our results provide important insight into key hazards and tradeoffs involved with revealing or not revealing gender identity and static or dynamic appearance when first or second impressions form and new relationships develop.

Our study provides an explanation for why cooperation is so commonly observed among strangers in social dilemmas like the PD despite incentives to not cooperate: people can mind-read, effectively intuiting the cooperation propensities of most other people in a way that should allow them to seek out and maintain mutually beneficial cooperative relationships. Prior studies demonstrated that cheater detection abilities are particularly sensitive to rule violation information (Brown & Moore, 2000; Cosmides, 1989b; Cosmides & Tooby, 1992; Fiddick & Erlich, 2010; Oda et al., 2006). Our study demonstrates that cheater and cooperator detection is sensitive to sparse person and context information, adding another facet to our understanding of how cheater detection adaptations are designed. Our study also provides insight into accurate predictions

PREDICTIVE MIND READING

of trust re-extension, an important but precarious and all-too-common problem in personal and business relationships (Robinson & Rousseau, 1994; Schniter & Sheremeta, 2014).

The behavioral sciences have closely studied the design of people's chosen behaviors in potentially cooperative strategic interactions. However, a clean experimental test and clear understanding of people's expectations of others' behaviors in unacquainted and repeated interactions have been largely missing. The evidence presented here suggest people can accurately predict the cooperativeness of strangers, helping explain the broad extent of human cooperativeness revealed by experimental and ethnographic studies. In conclusion, our study provides further support for the claim that an evolutionary–functional framework is a productive and promising approach to uncovering the nature of human mind reading.

PREDICTIVE MIND READING

Table 1. Rater demographics.

Treatment (N)	Guess response time (seconds per target):		Age in years	Self-identify with:			Prior familiarity with:		
	First Impression – Round 1	Second impression – Round 2		Male	Female	Other or prefer no answer	Player(s) in image	University students	University
None (108)	1.16 (1.05)	2.21 (1.49)	36.3 (12.7)	47.2	50.0	2.8	N/A	0.9 (9.6)	2.8 (16.5)
Label (101)	1.40 (0.64)	2.40 (1.44)	32.8 (11.3)	50.5	45.5	4.0	N/A	1.0 (9.9)	3.0 (17.1)
Photo (108)	3.13 (2.56)	3.70 (2.39)	33.6 (11.8)	49.1	50.0	0.9	0.9 (9.6)	0.9 (9.6)	0.9 (9.6)
Video (105)	8.32 (2.39)	8.82 (3.21)	32.6 (12.3)	48.6	49.5	1.9	0.0 (0.0)	0.0 (0.0)	1.9 (13.7)
Total (442)	3.50 (3.43)	4.28 (3.49)	33.9 (12.1)	48.8	48.8	3.6		0.7 (8.4)	2.1 (14.4)

Note: ‘Self-identify with’ values are percentages of raters’ responses. ‘Prior familiarity with’ values are percentages of raters with responses that indicate some familiarity (see online Appendix B for survey question details).

PREDICTIVE MIND READING

Table 2. Raters' stereotyped and generalized beliefs about PD players' cooperativeness.

	Stereotype about male players	Stereotype about female players	Generalized beliefs about all players
Belief revealed by:			
Male Rater N = 206	46.3 (18.6)	61.8 (18.4)	54.1 (16.6)
Female Rater N = 206	42.2 (17.7)	66.5 (15.2)	54.4 (13.8)
All Other Raters N = 10	40.3 (13.6)	50.4 (18.7)	45.4 (15.3)
Combined N = 422	44.2 (18.1)	63.9 (17.2)	54.0 (15.3)
Players' actual cooperativeness:			
Round 1	Males 61.4 (49.2)	Females 86.0 (35.1)	All players 74.5 (43.8)
Round 2	43.2 (50.1)	62.0 (49.0)	53.2 (50.2)
Both Rounds	52.3 (50.2)	74.0 (44.1)	63.8 (48.2)

Note. Where beliefs are reported, values are percent of time (SD in parenthesis) raters guess that each gender chooses 'Split' in round 1 of the repeated PD. Where players actual cooperativeness is reported, values are percent of time (SD in parenthesis) players choose 'Split'.

PREDICTIVE MIND READING

Table 3. Raters' guesses about PD players.

Panel A: Summary statistics

Treatment (N)	Guessed Split	Correctness	Cooperator Detection	Cheater Detection	Accuracy	Bias	Odds- ratio
<i>First Impression: Round 1 guesses</i> [actual Split = 74.5%]							
None (108)	63.6 (27.2)	56.1 (14.0)	63.2 (27.4)	35.4 (28.0)	-0.061 (0.324)	-0.480 (1.131)	0.990 (0.074)
Label (101)	61.0 (15.9)	58.7 (8.6)	63.2 (15.8)	45.5 (19.7)	0.238 (0.363)	-0.277 (0.548)	1.062 (0.092)
Photo (108)	56.1 (16.8)	55.9 (8.7)	58.1 (16.5)	49.6 (20.6)	0.214 (0.348)	-0.149 (0.619)	1.059 (0.095)
Video (105)	54.0 (13.2)	53.4 (8.5)	55.0 (13.9)	48.9 (15.9)	0.113 (0.362)	-0.087 (0.381)	1.028 (0.101)
All (422)	58.7 (19.4)	56.0 (10.4)	59.8 (19.5)	44.8 (22.2)	0.124 (0.368)	-0.249 (0.743)	1.034 (0.095)
<i>Second Impression: Round 2 guesses</i> [actual Split = 53.2%]							
None (108)	60.0 (18.4)	60.2 (8.2)	68.9 (20.4)	50.2 (19.7)	0.536 (0.451)	-0.268 (0.685)	1.188 (0.160)
Label (101)	60.8 (14.6)	60.7 (7.9)	70.3 (15.5)	49.9 (17.9)	0.561 (0.438)	-0.317 (0.511)	1.199 (0.159)
Photo (108)	59.0 (12.1)	62.5 (6.3)	70.2 (12.9)	53.8 (14.5)	0.674 (0.351)	-0.227 (0.361)	1.237 (0.126)
Video (105)	55.4 (11.1)	61.6 (6.5)	66.0 (12.9)	56.6 (12.9)	0.606 (0.352)	-0.131 (0.312)	1.223 (0.128)
All (422)	58.8 (14.5)	61.3 (7.3)	68.9 (15.8)	52.6 (16.6)	0.595 (0.402)	-0.235 (0.493)	1.212 (0.145)

Note: Values for guessed split, correctness, cooperator detection, and cheater detection are percentages. Standard deviations are in parentheses.

Panel B: Second-impression round 2 guess and correctness by players' behavioral history.

Behavioral History	Actual Split	Raters' round 2 guess				
		None	Label	Photo	Video	Combined
Both Take All	83.3	40.0	38.3	31.9	33.8	36.0
Take All/Partner Split	22.2	38.6	34.7	24.0	22.2	29.8
Split/Partner Take All	26.3	25.9	32.2	31.4	27.9	29.3
Both Split	70.6	82.6	83.4	84.8	79.9	82.7
Total	53.2	60.0	60.8	59.0	55.4	58.8
Behavioral History	Percentage of Players	Raters' round 2 correctness				
		None	Label	Photo	Video	Combined
Both Take All	6.4	43.4	41.1	39.4	41.0	41.2
Take All/Partner Split	19.1	55.7	59.8	66.0	65.6	61.8
Split/Partner Take All	20.2	61.0	58.8	60.2	61.2	60.3
Both Split	54.3	63.4	64.1	64.9	62.8	63.8
Total	100	60.2	60.7	62.5	61.6	61.3

Note: All values reported for raters' round 2 guess are percent of time raters guess players would choose Split. All values reported for raters' round 2 guess correctness are the percent of raters' round 2 guesses that are correct.

PREDICTIVE MIND READING

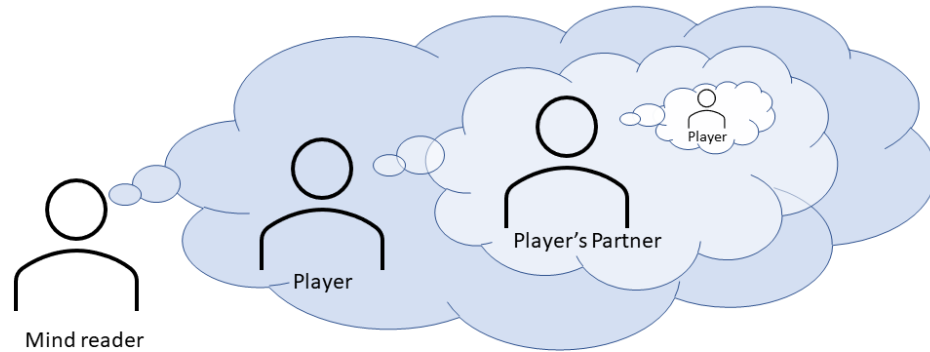
Table 4. Logit regression predicting Round 1 guesses controlling for the gender of PD players and raters' beliefs.

	(1)	(2)
Detected Gender	2.13 *** (5.59)	
Photo		0.52 *** (3.46)
Video		1.30 *** (8.78)
Stereotype	0.08 *** (13.66)	0.05 *** (30.47)
Detected Gender X Stereotype	-0.04 *** (-6.60)	
Photo X Stereotype		-0.01 *** (-5.04)
Video X Stereotype		-0.03 *** (-12.89)
Constant	-3.7 *** (-11.00)	-2.37 *** (-21.24)
Guess	39,668	29,516
Raters	422	314
Log-Likelihood	-24,031	-18,057
AIC	48,071	36,128
BIC	48,114	36,186
Chi-Squared (3 df) / (5 df)	323 ***	1,861 ***

T-statistic in parenthesis. ***: $p < .001$. Regression (1) includes round 1 data from all treatments. Detected gender is equal to one if the treatment is Label, Photo or Video. Regression (2) includes round 1 data only from treatments where player gender can be visually detected: Label, Photo, and Video. The variable *Stereotype* in regression (1) refers to raters' generalized beliefs (i.e., the average of their male and female stereotypes), whereas *Stereotype* in regression (2) refers to the applicable male or female stereotype given the player's self-description. As we control for Photo and Video in regression (2), *Stereotype*'s effect is from the Label treatment only.

PREDICTIVE MIND READING

To predict cooperation in any round of repeated Prisoner's Dilemma after round one, the mind reader accounts for goodwill that the player and their partner each demonstrated in the previous round and will next respond to.



Simple heuristic predicting cooperation: The player will cooperate in the next round only if mutual cooperation occurred in the previous round. If either the player or their partner did not cooperate, the player will not cooperate in the next round.

Figure 1. Mind readers' goodwill accounting heuristic for predicting players' cooperative propensity in a repeated prisoner's dilemma game with unknown endgame.

PREDICTIVE MIND READING

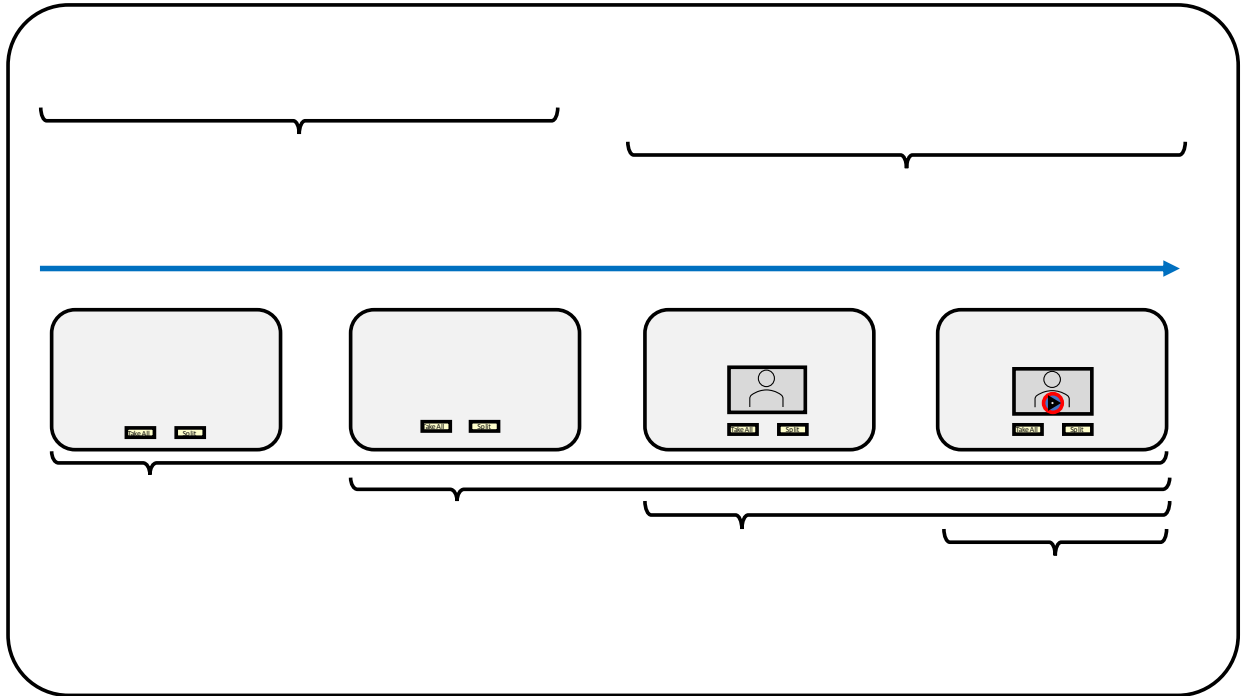


Figure 2. Schematic diagram showing incremental manipulation of gender, face and static appearance, dynamic appearance, and behavioral history information available for first and second impressions in a predictive mind reading experiment.

PREDICTIVE MIND READING

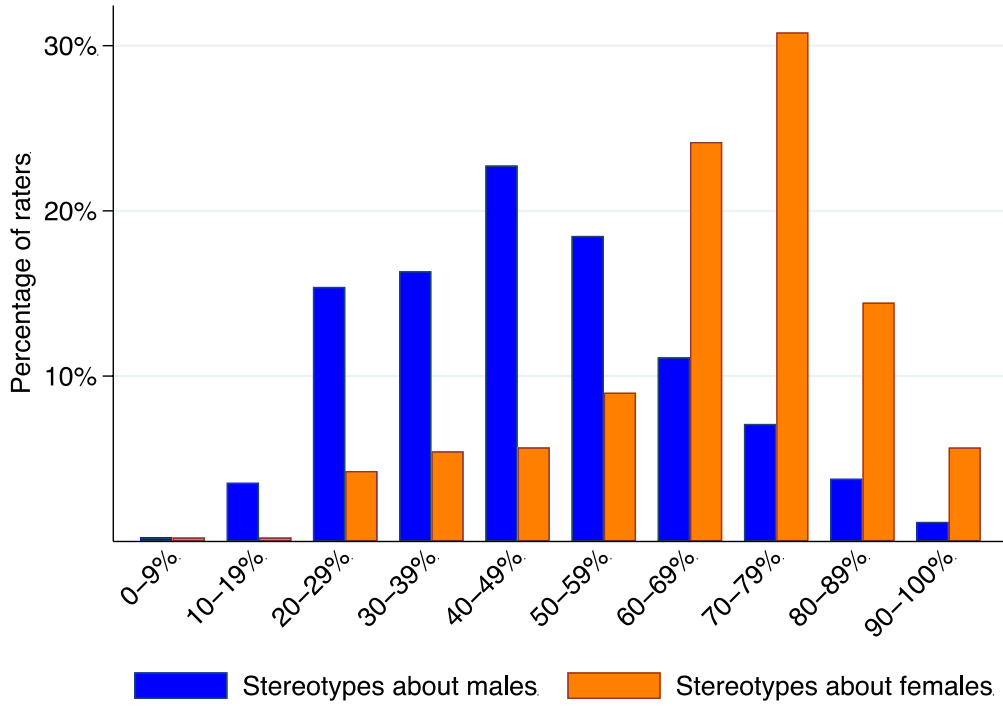


Figure 3. Raters' gender-specific stereotype beliefs about male and female players of a repeated prisoner's dilemma game with unknown endgame.

PREDICTIVE MIND READING

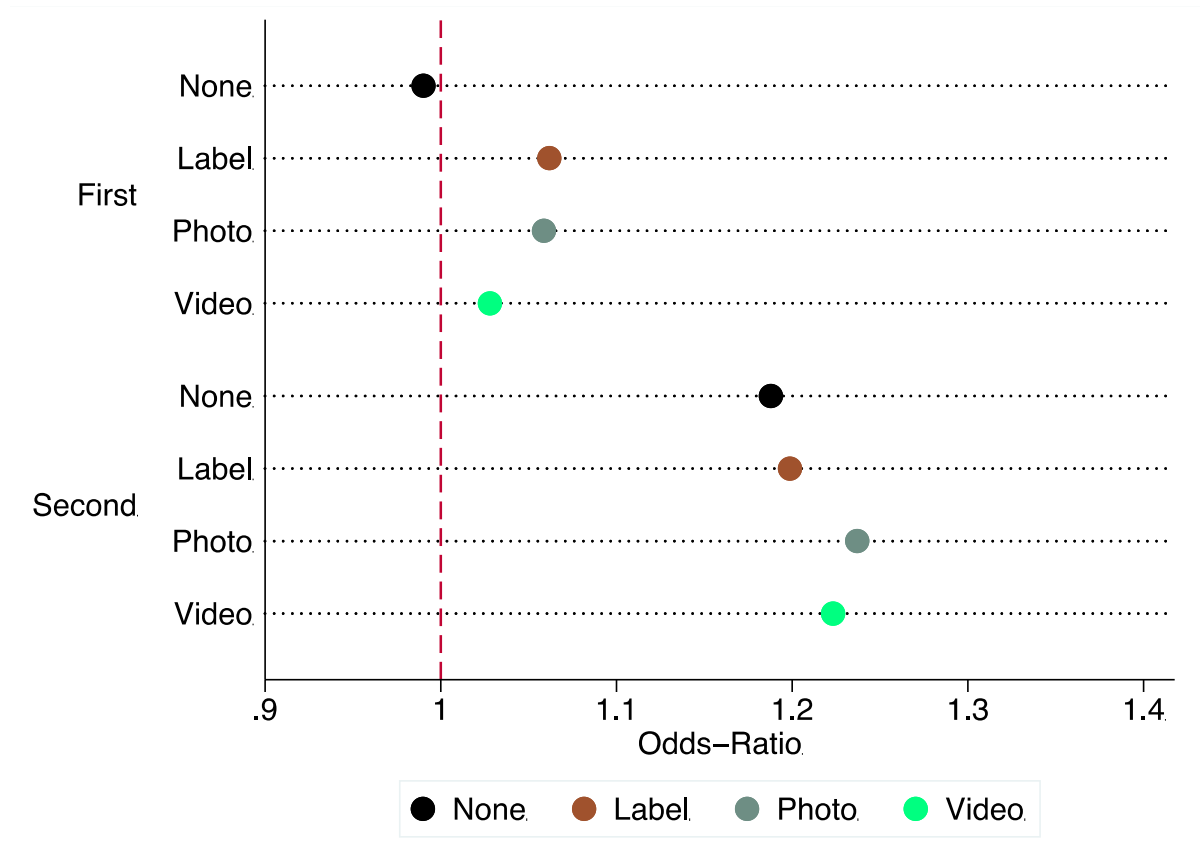


Figure 4. Odds-ratio of guesses by first and second impression. The odds ratio is the number of correct guesses in a round divided by the expected number of correct guesses due to chance alone. Better-than-chance guesses are defined by values greater than one. The maximum odds-ratio value, for a rater that guesses correctly 100% of the time, is 1.61 and 1.99 for first and second round guesses, respectively.

References

- Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the Enigmatic Face: The Importance of Facial Dynamics in Interpreting Subtle Facial Expressions. *Psychological Science*, *16*(5), 403–410. <https://doi.org/10.1111/j.0956-7976.2005.01548.x>
- Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In *Advances in Experimental Social Psychology* (Vol. 32, pp. 201–271). Academic Press. [https://doi.org/10.1016/S0065-2601\(00\)80006-4](https://doi.org/10.1016/S0065-2601(00)80006-4)
- Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, *64*(3), 431–441. <https://doi.org/10.1037/0022-3514.64.3.431>
- Ames, D. R., Weber, E. U., & Zou, X. (2012). Mind-reading in strategic interaction: The impact of perceived similarity on projection and stereotyping. *Organizational Behavior and Human Decision Processes*, *117*(1), 96–110. <https://doi.org/10.1016/j.obhdp.2011.07.007>
- Andreoni, J., & Miller, J. H. (1993). Rational Cooperation in the Finitely Repeated Prisoner’s Dilemma: Experimental Evidence. *The Economic Journal*, *103*(418), 570–585. <https://doi.org/10.2307/2234532>
- Andreoni, J., & Petrie, R. (2008). Beauty, gender and stereotypes: Evidence from laboratory experiments. *Journal of Economic Psychology*, *29*(1), 73–93. <https://doi.org/10.1016/j.joep.2007.07.008>
- Balliet, D., & Van Lange, P. A. M. (2013). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*, *139*(5), 1090–1112. <https://doi.org/10.1037/a0030939>
- Barclay, P. (2008). Enhanced recognition of defectors depends on their rarity. *Cognition*, *107*(3), 817–828. <https://doi.org/10.1016/j.cognition.2007.11.013>
- Baron, R. S. (1986). Distraction-Conflict Theory: Progress and Problems. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 19, pp. 1–40). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60211-7](https://doi.org/10.1016/S0065-2601(08)60211-7)
- Baron-Cohen, S. (1997). *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press.
- Blair, I. V., Judd, C. M., & Fallman, J. L. (2004). The Automaticity of Race and Afrocentric Facial Features in Social Judgments. *Journal of Personality and Social Psychology*, *87*(6), 763–778. <https://doi.org/10.1037/0022-3514.87.6.763>
- Bó, P. D. (2005). Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games. *American Economic Review*, *95*(5), 1591–1604. <https://doi.org/10.1257/000282805775014434>
- Bonnefon, J.-F., Hopfensitz, A., & De Neys, W. (2013). The modular nature of trustworthiness detection. *Journal of Experimental Psychology: General*, *142*(1), 143–150. <https://doi.org/10.1037/a0028930>
- Brosig, J. (2002). Identifying cooperative behavior: Some experimental results in a prisoner’s dilemma game. *Journal of Economic Behavior & Organization*, *47*(3), 275–290. [https://doi.org/10.1016/S0167-2681\(01\)00211-6](https://doi.org/10.1016/S0167-2681(01)00211-6)
- Brown, W. M., & Moore, C. (2000). Is prospective altruist-detection an evolved solution to the adaptive problem of subtle cheating in cooperative ventures? Supportive evidence using the Wason selection task. *Evolution and Human Behavior*, *21*(1), 25–37. [https://doi.org/10.1016/S1090-5138\(99\)00018-5](https://doi.org/10.1016/S1090-5138(99)00018-5)
- Brown, W. M., Palameta, B., & Moore, C. (2003). Are there Nonverbal Cues to Commitment? An Exploratory Study Using the Zero-Acquaintance Video Presentation Paradigm. *Evolutionary Psychology*, *1*(1), 147470490300100100. <https://doi.org/10.1177/147470490300100104>
- Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., & Linney, A. (1993). Sex Discrimination: How Do We Tell the Difference between Male and Female Faces? *Perception*, *22*(2), 131–152. <https://doi.org/10.1068/p220131>

PREDICTIVE MIND READING

- Bruce, V., & Young, A. (2011). *Face Perception*. Psychology Press.
<https://doi.org/10.4324/9780203721254>
- Burton-Chellew, M. N., & West, S. A. (2012). Correlates of Cooperation in a One-Shot High-Stakes Televised Prisoners' Dilemma. *PLOS ONE*, 7(4), e33344.
<https://doi.org/10.1371/journal.pone.0033344>
- Camerer, C., & Weigelt, K. (1998). Experimental Tests of a Sequential Equilibrium Reputation Model. *Econometrica*, 56, 1–36.
- Coricelli, G., McCabe, K., & Smith, V. (2000). Theory-of-mind mechanism in personal exchange. In *Hatano, et al. (Eds.), Affective minds* (pp. 249–259).
- Cosmides, L. (1989a). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31(3), 187–276. [https://doi.org/10.1016/0010-0277\(89\)90023-1](https://doi.org/10.1016/0010-0277(89)90023-1)
- Cosmides, L. (1989b). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31(3), 187–276. [https://doi.org/10.1016/0010-0277\(89\)90023-1](https://doi.org/10.1016/0010-0277(89)90023-1)
- Cosmides, L., & Tooby, J. (1989). Evolutionary psychology and the generation of culture, part II: Case study: A computational theory of social exchange. *Ethology and Sociobiology*, 10(1), 51–97.
[https://doi.org/10.1016/0162-3095\(89\)90013-7](https://doi.org/10.1016/0162-3095(89)90013-7)
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. *The Adapted Mind : Evolutionary Psychology and the Generation of Culture*, 163–228.
- Cosmides, L., & Tooby, J. (2005). Neurocognitive adaptations designed for social exchange. *The Handbook of Evolutionary Psychology*, 584–627.
- Dal Bó, P., & Fréchette, G. R. (2011). The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence. *American Economic Review*, 101(1), 411–429.
<https://doi.org/10.1257/aer.101.1.411>
- Daly, M., & Wilson, M. (1988). *Homicide*. Transaction Publishers.
- Dawes, R. M., & Thaler, R. H. (1988). Anomalies: Cooperation. *Journal of Economic Perspectives*, 2(3), 187–197. <https://doi.org/10.1257/jep.2.3.187>
- Dennett, D. C. (1989). *The Intentional Stance*. MIT Press.
- DeSteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L., & Lee, J. J. (2012). Detecting the Trustworthiness of Novel Partners in Economic Exchange. *Psychological Science*, 23(12), 1549–1556. <https://doi.org/10.1177/0956797612448793>
- Dickhaut, J., Hubbard, J., McCabe, K., & Smith, V. (2008). *Trust, reciprocity, and interpersonal history: Fool me once, shame on you, fool me twice, shame on me*. Working paper, University of Minnesota.
- Durkin, M. P., Jollineau, S. J., & Lyon, S. C. (2020). Sounds Good to Me: How Communication Mode and Priming Affect Auditor Performance. *AUDITING: A Journal of Practice & Theory*, 40(1), 1–17. <https://doi.org/10.2308/AJPT-19-038>
- Eagly, A. H. (2009). The his and hers of prosocial behavior: An examination of the social psychology of gender. *American Psychologist*, 64(8), 644–658. <https://doi.org/10.1037/0003-066X.64.8.644>
- Efferson, C., & Vogt, S. (2013). Viewing men's faces does not lead to accurate predictions of trustworthiness. *Scientific Reports*, 3(1), 1047. <https://doi.org/10.1038/srep01047>
- Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., Tomita, M., & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53(4), 712–717. <https://doi.org/10.1037/0022-3514.53.4.712>
- Embrey, M., Fréchette, G. R., & Yuksel, S. (2018). Cooperation in the Finitely Repeated Prisoner's Dilemma*. *The Quarterly Journal of Economics*, 133(1), 509–551.
<https://doi.org/10.1093/qje/qjx033>

PREDICTIVE MIND READING

- Fasolt, V., Holzleitner, I. J., Lee, A. J., O'Shea, K. J., & DeBruine, L. M. (2019). Contribution of shape and surface reflectance information to kinship detection in 3D face images. *Journal of Vision*, *19*(12), 9–9. <https://doi.org/10.1167/19.12.9>
- Fehr, E., & Henrich, J. (2003). Is Strong Reciprocity a Maladaptation? On the Evolutionary Foundations of Human Altruism. In *Genetic and Cultural Evolution of Cooperation*, ed P. Hammerstein (pp. 55–82). MIT Press. <https://doi.org/10.2139/ssrn.382950>
- Fetchenhauer, D., & Dunning, D. (2010). Why So Cynical?: Asymmetric Feedback Underlies Misguided Skepticism Regarding the Trustworthiness of Others. *Psychological Science*, *21*(2), 189–193. <https://doi.org/10.1177/0956797609358586>
- Fetchenhauer, D., Groothuis, T., & Pradel, J. (2010). Not only states but traits—Humans can identify permanent altruistic dispositions in 20 s. *Evolution and Human Behavior*, *31*(2), 80–86. <https://doi.org/10.1016/j.evolhumbehav.2009.06.009>
- Fiddick, L., & Erlich, N. (2010). Giving it all away: Altruism and answers to the Wason selection task. *Evolution and Human Behavior*, *31*(2), 131–140. <https://doi.org/10.1016/j.evolhumbehav.2009.08.003>
- Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, *37*(4), 571–583.
- Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions* (pp. xiii, 304). W W Norton & Co.
- Frank, R. H., Gilovich, T., & Regan, D. T. (1993). The evolution of one-shot cooperation: An experiment. *Ethology and Sociobiology*, *14*(4), 247–256. [https://doi.org/10.1016/0162-3095\(93\)90020-I](https://doi.org/10.1016/0162-3095(93)90020-I)
- Frith, C. D., & Frith, U. (2006). The Neural Basis of Mentalizing. *Neuron*, *50*(4), 531–534. <https://doi.org/10.1016/j.neuron.2006.05.001>
- Funder, D. C. (2012). Accurate Personality Judgment. *Current Directions in Psychological Science*, *21*(3), 177–182. <https://doi.org/10.1177/0963721412445309>
- Geniole, S. N., Denson, T. F., Dixson, B. J., Carré, J. M., & McCormick, C. M. (2015). Evidence from Meta-Analyses of the Facial Width-to-Height Ratio as an Evolved Cue of Threat. *PLOS ONE*, *10*(7), e0132726. <https://doi.org/10.1371/journal.pone.0132726>
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669. <https://doi.org/10.1037/0033-295X.103.4.650>
- Green, M. J., & Phillips, M. L. (2004). Social threat perception and the evolution of paranoia. *Neuroscience & Biobehavioral Reviews*, *28*(3), 333–342. <https://doi.org/10.1016/j.neubiorev.2004.03.006>
- Harwood, N. K., Hall, L. J., & Shinkfield, A. J. (1999). Recognition of Facial Emotional Expressions From Moving and Static Displays by Individuals with Mental Retardation. *American Journal on Mental Retardation*, *104*(3), 270–278. [https://doi.org/10.1352/0895-8017\(1999\)104<0270:ROFEEF>2.0.CO;2](https://doi.org/10.1352/0895-8017(1999)104<0270:ROFEEF>2.0.CO;2)
- Haselton, M. G., & Nettle, D. (2006). The Paranoid Optimist: An Integrative Evolutionary Model of Cognitive Biases. *Personality and Social Psychology Review*, *10*(1), 47–66.
- Haselton, M. G., Nettle, D., & Murray, D. R. (2015). The Evolution of Cognitive Bias. In *The Handbook of Evolutionary Psychology* (pp. 1–20). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119125563.evpsych241>
- Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, *78*(5), 837–852. <https://doi.org/10.1037/0022-3514.78.5.837>
- Hayashi, N., Ostrom, E., Walker, J., & Yamagishi, T. (1999). RECIPROCITY, TRUST, AND THE SENSE OF CONTROL: A CROSS-SOCIETAL STUDY. *Rationality and Society*, *11*(1), 27–46. <https://doi.org/10.1177/104346399011001002>

- Hill, H., Bruce, V., & Akamatsu, S. (1995). Perceiving the sex and race of faces: The role of shape and colour. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 261(1362), 367–373. <https://doi.org/10.1098/rspb.1995.0161>
- Hirshleifer, J. (1987). On the emotions as guarantors of threats and promises. *The Latest on the Best: Essays on Evolution and Optimality*, 307–326.
- Jaeger, B., Oud, B., Williams, T., Krumhuber, E. G., Fehr, E., & Engelmann, J. B. (2022). Can people detect the trustworthiness of strangers based on their facial appearance? *Evolution and Human Behavior*. <https://doi.org/10.1016/j.evolhumbehav.2022.04.004>
- Jaeger, B., Slegers, W. W. A., & Evans, A. M. (2020). Automated classification of demographics from face images: A tutorial and validation. *Social and Personality Psychology Compass*, 14(3), e12520. <https://doi.org/10.1111/spc3.12520>
- Jussim, L., Cain, T. R., Crawford, J. T., Harber, K., & Cohen, F. (2009). The unbearable accuracy of stereotypes. In *Handbook of prejudice, stereotyping, and discrimination* (pp. 199–227). Psychology Press.
- Kaplan, H. S., Schniter, E., Smith, V. L., & Wilson, B. J. (2012). Risk and the evolution of human exchange. *Proceedings of the Royal Society B: Biological Sciences*, 279(1740), 2930–2935. <https://doi.org/10.1098/rspb.2011.2614>
- Kaplan, H. S., Schniter, E., Smith, V. L., & Wilson, B. J. (2018). Experimental tests of the tolerated theft and risk-reduction theories of resource exchange. *Nature Human Behaviour*, 2(6), 383. <https://doi.org/10.1038/s41562-018-0356-x>
- Kiyonari, T. (2010). Detecting Defectors When They Have Incentives to Manipulate Their Impressions. *Letters on Evolutionary Behavioral Science*, 1(1), Article 1. <https://doi.org/10.5178/lebs.2010.5>
- Kiyonari, T., Tanida, S., & Yamagishi, T. (2000). Social exchange and reciprocity: Confusion or a heuristic? *Evolution and Human Behavior*, 21(6), 411–427. [https://doi.org/10.1016/S1090-5138\(00\)00055-6](https://doi.org/10.1016/S1090-5138(00)00055-6)
- Kreps, D. M., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, 27(2), 245–252. [https://doi.org/10.1016/0022-0531\(82\)90029-1](https://doi.org/10.1016/0022-0531(82)90029-1)
- Kurzban, R., & Houser, D. (2005). Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *Proceedings of the National Academy of Sciences*, 102(5), 1803–1807. <https://doi.org/10.1073/pnas.0408759102>
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 21(3), 451–468. <https://doi.org/10.1037/0096-1523.21.3.451>
- Löckenhoff, C. E., Chan, W., McCrae, R. R., De Fruyt, F., Jussim, L., De Bolle, M., Costa, P. T., Sutin, A. R., Realo, A., Allik, J., Nakazato, K., Shimonaka, Y., Hřebíčková, M., Graf, S., Yik, M., Ficková, E., Brunner-Sciarrà, M., Leibovich de Figueora, N., Schmidt, V., ... Terracciano, A. (2014). Gender Stereotypes of Personality: Universal and Accurate? *Journal of Cross-Cultural Psychology*, 45(5), 675–694. <https://doi.org/10.1177/0022022113520075>
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection Theory: A User's Guide*. Taylor & Francis Group.
- Manson, J. H., Gervais, M. M., & Kline, M. A. (2013). Defectors Cannot Be Detected during “Small Talk” with Strangers. *PLOS ONE*, 8(12), e82531. <https://doi.org/10.1371/journal.pone.0082531>
- Mao, A., Dworkin, L., Suri, S., & Watts, D. J. (2017). Resilient cooperators stabilize long-run cooperation in the finitely repeated Prisoner's Dilemma. *Nature Communications*, 8(1), 13800. <https://doi.org/10.1038/ncomms13800>
- Martin, D. L., & Frayer, David. W. (Eds.). (2014). *Troubled Times: Violence and Warfare in the Past*. Routledge. <https://doi.org/10.4324/9781315078328>
- McCabe, K. A., Rassenti, S. J., & Smith, V. L. (1996). Game theory and reciprocity in some extensive form experimental games. *Proceedings of the National Academy of Sciences*, 93(23), 13421–13428. <https://doi.org/10.1073/pnas.93.23.13421>

PREDICTIVE MIND READING

- McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization*, 52(2), 267–275. [https://doi.org/10.1016/S0167-2681\(03\)00003-9](https://doi.org/10.1016/S0167-2681(03)00003-9)
- McCabe, K. A., & Smith, V. L. (2001). Goodwill accounting and the process of exchange. *Bounded Rationality: The Adaptive Toolbox*, 319–340.
- McCabe, K. A., Smith, V. L., & LePore, M. (2000). Intentionality detection and “mindreading”: Why does game form matter? *Proceedings of the National Academy of Sciences*, 97(8), 4404–4409. <https://doi.org/10.1073/pnas.97.8.4404>
- McKelvey, R. D., & Palfrey, T. R. (1992). An Experimental Study of the Centipede Game. *Econometrica*, 60(4), 803–836. <https://doi.org/10.2307/2951567>
- Nesse, R. M. (2005). Natural selection and the regulation of defenses: A signal detection analysis of the smoke detector principle. *Evolution and Human Behavior*, 26(1), 88–105. <https://doi.org/10.1016/j.evolhumbehav.2004.08.002>
- Normann, H.-T., & Wallace, B. (2012). The impact of the termination rule on cooperation in a prisoner’s dilemma experiment. *International Journal of Game Theory*, 41(3), 707–718. <https://doi.org/10.1007/s00182-012-0341-y>
- Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437(7063), 1291–1298. <https://doi.org/10.1038/nature04131>
- Oda, R., Hiraishi, K., & Matsumoto-Oda, A. (2006). Does an altruist-detection cognitive mechanism function independently of a cheater-detection cognitive mechanism? Studies using Wason selection tasks. *Evolution and Human Behavior*, 27(5), 366–380. <https://doi.org/10.1016/j.evolhumbehav.2006.03.002>
- Oda, R., Yamagata, N., Yabiku, Y., & Matsumoto-Oda, A. (2009). Altruism Can Be Assessed Correctly Based on Impression. *Human Nature*, 20(3), 331–341. <https://doi.org/10.1007/s12110-009-9070-8>
- Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, 46(2), 315–324. <https://doi.org/10.1016/j.jesp.2009.12.002>
- Ostrom, E., & Walker, J. (2003). *Trust and Reciprocity: Interdisciplinary Lessons for Experimental Research*. Russell Sage Foundation.
- Palermo, R., & Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia*, 45(1), 75–92. <https://doi.org/10.1016/j.neuropsychologia.2006.04.025>
- Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences*, 99(17), 11458–11463. <https://doi.org/10.1073/pnas.172403899>
- Pike, G. E., Kemp, R. I., Towell, N. A., & Phillips, K. C. (1997). Recognizing Moving Faces: The Relative Contribution of Motion and Perspective View Information. *Visual Cognition*, 4(4), 409–438. <https://doi.org/10.1080/713756769>
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4), 515–526. <https://doi.org/10.1017/S0140525X00076512>
- Reed, L. I., Zeglen, K. N., & Schmidt, K. L. (2012a). Facial expressions as honest signals of cooperative intent in a one-shot anonymous Prisoner’s Dilemma game. *Evolution and Human Behavior*, 33(3), 200–209. <https://doi.org/10.1016/j.evolhumbehav.2011.09.003>
- Reed, L. I., Zeglen, K. N., & Schmidt, K. L. (2012b). Facial expressions as honest signals of cooperative intent in a one-shot anonymous Prisoner’s Dilemma game. *Evolution and Human Behavior*, 33(3), 200–209. <https://doi.org/10.1016/j.evolhumbehav.2011.09.003>
- Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable Facial Configurations Affect Strategic Choices in Trust Games with or without Information about Past Behavior. *PLOS ONE*, 7(3), e34293. <https://doi.org/10.1371/journal.pone.0034293>

PREDICTIVE MIND READING

- Robinson, S. L., & Rousseau, D. M. (1994). Violating the psychological contract: Not the exception but the norm. *Journal of Organizational Behavior*, *15*(3), 245–259.
<https://doi.org/10.1002/job.4030150306>
- Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., & Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: An fMRI study. *Cognitive Brain Research*, *20*(1), 81–91. <https://doi.org/10.1016/j.cogbrainres.2004.01.008>
- Schaller, M. (2008). Evolutionary bases of first impressions. In *First Impressions*. Edited by Nalini Ambady and John Joseph Skowronski. (pp. 15–34). The Guilford Press.
- Schneider, M., & Shields, T. (2022). Motives for Cooperation in the One-Shot Prisoner's Dilemma. *Journal of Behavioral Finance*, *0*(0), 1–19. <https://doi.org/10.1080/15427560.2022.2081974>
- Schniter, E., & Sheremeta, R. M. (2014). Predictable and predictive emotions: Explaining cheap signals and trust re-extension. *Frontiers in Behavioral Neuroscience*, *8*.
<https://doi.org/10.3389/fnbeh.2014.00401>
- Schniter, E., & Shields, T. W. (2014). Ageism, honesty, and trust. *Journal of Behavioral and Experimental Economics*, *51*, 19–29. <https://doi.org/10.1016/j.socec.2014.03.006>
- Schniter, E., & Shields, T. W. (2020). Gender, Stereotypes, and Trust in Communication. *Human Nature*, *31*(3), 296–321. <https://doi.org/10.1007/s12110-020-09376-3>
- Schug, J., Matsumoto, D., Horita, Y., Yamagishi, T., & Bonnet, K. (2010). Emotional expressivity as a signal of cooperation. *Evolution and Human Behavior*, *31*(2), 87–94.
<https://doi.org/10.1016/j.evolhumbehav.2009.09.006>
- Seabright, P. (2010). *The Company of Strangers: A Natural History of Economic Life - Revised Edition*. Princeton University Press.
- Sell, A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., & Gurven, M. (2009). Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1656), 575–584.
<https://doi.org/10.1098/rspb.2008.1177>
- Sell, A., Hone, L. S. E., & Pound, N. (2012). The Importance of Physical Strength to Human Males. *Human Nature*, *23*(1), 30–44. <https://doi.org/10.1007/s12110-012-9131-2>
- Snyder, J. A., Tabler, J., & Gonzales, C. M. (2022). Measuring Sexual Identity, Gender Identity, and Biological Sex in Large Social Surveys: Implications for Victimization Research. *Criminal Justice and Behavior*, *49*(9), 1376–1395. <https://doi.org/10.1177/00938548221097034>
- Snyder, M. (1984). When Belief Creates Reality. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 18, pp. 247–305). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60146-X](https://doi.org/10.1016/S0065-2601(08)60146-X)
- Sparks, A., Burleigh, T., & Barclay, P. (2016). We can see inside: Accurate prediction of Prisoner's Dilemma decisions in announced games following a face-to-face interaction. *Evolution and Human Behavior*, *37*(3), 210–216. <https://doi.org/10.1016/j.evolhumbehav.2015.11.003>
- Sylwester, K., Lyons, M., Buchanan, C., Nettle, D., & Roberts, G. (2012). The role of Theory of Mind in assessing cooperative intentions. *Personality and Individual Differences*, *52*(2), 113–117.
<https://doi.org/10.1016/j.paid.2011.09.005>
- Todd, P. M. (2001). Fast and frugal heuristics for environmentally bounded minds. In *Bounded rationality: The adaptive toolbox* (pp. 51–70). The MIT Press.
- Todorov, A. (2017). Face Value: The Irresistible Influence of First Impressions. In *Face Value*. Princeton University Press. <https://doi.org/10.1515/9781400885725>
- Tognetti, A., Berticat, C., Raymond, M., & Faurie, C. (2013). Is cooperativeness readable in static facial features? An inter-cultural approach. *Evolution and Human Behavior*, *34*(6), 427–432.
<https://doi.org/10.1016/j.evolhumbehav.2013.08.002>
- van den Assem, M. J., van Dolder, D., & Thaler, R. H. (2012). Split or Steal? Cooperative Behavior When the Stakes Are Large. *Management Science*, *58*(1), 2–20.
<https://doi.org/10.1287/mnsc.1110.1413>

PREDICTIVE MIND READING

- Verplaetse, J., Vanneste, S., & Braeckman, J. (2007). You can judge a book by its cover: The sequel.: A kernel of truth in predictive cheating detection. *Evolution and Human Behavior*, 28(4), 260–271. <https://doi.org/10.1016/j.evolhumbehav.2007.04.006>
- Vogt, S., Efferson, C., & Fehr, E. (2013). Can we see inside? Predicting strategic behavior given limited information. *Evolution and Human Behavior*, 34(4), 258–264. <https://doi.org/10.1016/j.evolhumbehav.2013.03.003>
- Whiten, A., & Byrne, R. W. (1991). *Natural theories of mind: Evolution, development and simulation of everyday mindreading*. B. Blackwell Oxford, UK.
- Zilioli, S., Sell, A. N., Stirrat, M., Jagore, J., Vickerman, W., & Watson, N. V. (2015). Face of a fighter: Bizygomatic width as a cue of formidability. *Aggressive Behavior*, 41(4), 322–330. <https://doi.org/10.1002/ab.21544>

PREDICTIVE MIND READING

Online Supplementary Materials for “Predictive Mind Reading from First and Second Impressions”.

Appendix A

Table A1: Correct guess of player behavior conditioned on behavioral history, sufficiently correct median, and PD player gender

Both Take All	-0.69 *** (-5.38)
Take All/Partner Split	0.16 (1.91)
Split/Partner Take All	0.02 (0.29)
Both Split	0.56 *** (10.94)
Sufficiently correct median	0.31 *** (10.19)
Sufficiently correct median X Both Take All	-0.31 ** (-2.61)
Sufficiently correct median X Take All/Partner Split	-0.26 *** (-3.44)
Sufficiently correct median X Split/Partner Take All	-0.27 *** (-3.66)
Sufficiently correct median X Both Split	-0.19 *** (-3.61)
Male player	-0.43 *** (-15.77)
Male player X Both Take All	0.50 *** (4.78)
Male player X Take All/Partner Split	0.60 *** (8.91)
Male player X Split/Partner Take All	0.78 *** (12.11)
Male player X Both Split	-0.33 *** (-7.65)
Constant	0.25 *** (7.69)
N	59,032
Groups	314
Log-Likelihood	-38,938.03
AIC	77,908.05
BIC	78,051.83
Chi-Sq (14 df)	1,800.65 ***

T-statistic in parenthesis. **: $p < .01$, ***: $p < .001$. The baseline history is the no history condition (first impression). Regression is on those treatments where the player gender could be visually detected.

PREDICTIVE MIND READING

Table A2: Round 2 guesses and correctness conditioned on behavioral history and treatments with photos and videos

	Guess	Correct
Both Take All	-2.45 *** (-24.55)	-0.83 *** (-9.79)
Take All/Partner Split	-2.52 *** (-36.66)	-0.33 *** (-6.07)
Split/Partner Take All	-3.23 *** (-44.02)	-0.10 (-1.93)
Faces	-0.15 (-1.23)	0.02 (0.52)
Both Take All X Faces	-0.03 (-0.26)	-0.14 (-1.44)
Take All/Partner Split X Faces	-0.38 *** (-4.73)	0.33 *** (5.17)
Split/Partner Take All X Faces	0.54 *** (6.58)	-0.06 (-0.99)
Constant	1.91 *** (18.37)	0.56 *** (15.63)
N	39,668	39,668
Groups	422	422
Log-Likelihood	-19,176.8	-26,140.2
AIC	38,371.6	52,298.3
BIC	38,448.8	52,375.6
Chi-Sq (7 df)	9,817.0 ***	514.3 ***

T-statistic in parenthesis. ***: $p < .001$. the condition ‘Both Split’ is the baseline. Faces is equal to one if the treatment is photo or video.

PREDICTIVE MIND READING

Appendix B

Experimental Task Instructions and Survey Questions

A unique set of questions is being generated that will test your eligibility to complete this study. Participants who answer these questions incorrectly are NOT eligible, will be screened out immediately, and will NOT receive any payment.

What did you see?

[One of the images below was randomly selected and displayed]

- A blue rectangle and black oval
- A green cross and orange triangle
- A red circle and green star
- A yellow triangle and red circle
- A black oval and blue cross
- A pink square and purple star

What is the answer to this question?

[One of four randomly selected word problems, with unique answer below, would appear here]

- Six
- Five
- Twelve
- Ten

Consent

We invite you to participate in a research study being conducted by XXXXX and XXXXX, both professors from XXXXX. The purpose of the study is
If you agree to participate, we would like you to complete tasks that involve making guesses and answering survey questions. The study will take on average 30 minutes to complete. Your payment will be your wage and a bonus. The bonus will be as high as \$3 depending on the accuracy of your predictions. At the end of the study, you will learn your task performance and the associated payment amount. You will receive payment for completing the study in its entirety. There is minimal foreseeable risk associated with this study. All responses are anonymous. Taking part in this research study is completely voluntary. If you do NOT wish to participate in this study, you can exit the study anytime. However, incomplete responses cannot be used for research and therefore, you will NOT receive payment. If you have any questions about the study, please contact XXXXX. If you have any questions about your rights as a research participant, please contact the Human Subjects Office at XXXXX. Thank you very much for your consideration of this research study. Select the appropriate option below to indicate whether you agree to participate.

- Yes, I agree to participate in this study
- No, I do NOT agree to participate in this study

Introduction

This survey is part of an experiment in the economics of decision-making. Various research agencies have provided funds for this research. By following the instructions carefully and making good decisions, you may earn an additional amount of money besides the payment for

PREDICTIVE MIND READING

completing the survey. The actual amount of additional money that you may earn will depend on your decisions. Your earnings will be reported to you after you complete the survey.

There are a couple of basic rules you must follow:

- 1) Do not communicate with others, or allow others to assist you while completing the survey.
- 2) Do not discuss this survey with others after completing the survey.
- 3) If you have any questions, comments, or concerns please communicate them to the researchers running this study by xxxx.

General Description of the Survey

The survey is broken into four separate parts. All four parts must be completed to earn payment. After you finish you will be provided with a Completion ID. In the first, second, and third parts of the survey, one randomly chosen guess determines your additional earnings from that part. Your overall income from the survey will be based on the sum of earnings from the first three parts of the survey and from your wage for completing all sections of the survey. It is in your best interest to make a careful decision in all possible situations. Researchers at XXX have previously conducted an experiment using an anonymous economic interaction (between a randomly paired Person 1 and Person 2) over a computer network. Participants in the original

x
you to make guesses about what participants did in that original experiment. A description of that experiment follows.

IMPORTANT: You will NOT participate in the experiment explained below, but it is important that you understand it because you will make guesses related to people and decisions from this original experiment conducted at XXX. Below are the instructions that were provided to participants in that original experiment.

The First-Round Decision [hover over text]

In this experiment you will be randomly paired with one other person. The other person, like yourself, was recruited from the XXX. During this experiment you will be paired with this person through a computer network. At no time will your true identity be revealed to the other participants here today. Even though we are video recording for research purposes, during this experiment your video recorded image will never be transmitted or shown to other participants

, x

The choices made by both you and the other person will affect how a \$10 dollar jackpot will be allocated. Because these choices affect your earnings, you should make a deliberate and conscious choice. You can either choose *Split* or you can choose to *Take All*. The other person you are paired with makes the same choice. If BOTH you and the person you are paired with choose *Split* you will both get a payoff of \$5. If you BOTH choose to *Take All* you will both get a payoff of \$0. If you choose to *Take All* but the other person chooses *Split* you will get a payoff of \$10 and the other person will receive \$0. Likewise, if you choose *Split* but the other person chooses to *Take All*, then you receive \$0 and the other person receives \$10. These payoffs are summarized in the table below. The bold number in each quadrant of the box below is the payment received by you, the other non-bold number is the payment received by the other person:

		The Other Person Chooses	
		<i>Split</i>	<i>Take All</i>
You Choose	<i>Split</i>	Your payoff 5 Other's payoff 5	Your payoff 0 Other's payoff 10
	<i>Take All</i>	Your payoff 10 Other's payoff 0	Your payoff 0 Other's payoff 0

When making your choice, you will not know the choice made by the other person. We first will ask you to state your intention while being recorded. That is, we want you to make a non-binding statement about what you intend to do, *Split* or *Take All*. Next, we ask you to make your choice: *Split* or *Take All*. After everyone in the experiment has made their choice, the computer will report the results: your choice, the choice made by the other person, and your payoffs. After reviewing the results, you will be asked to complete a short survey before moving on.

Set of Multiple Rounds

You will participate in a set composed of more than a single round. Each round is the same. The number of rounds that you will participate in is determined probabilistically by the computer. After each round has been finished, the probability of participating in another round is $(1/4)^{(n-1)}$ where n is the number of rounds so far. Since $(1/4)^0 = 1$ there will be at least 2 rounds and some probability of future rounds. However, it is uncertain how many more rounds there will be beyond these first 2 rounds. In all rounds you will interact with the same person that you were paired with in the first round.

Reminder

You will be participating in a set of two or more rounds and interacting with the same other person for all rounds. Even though we are video recording for research purposes, you are guaranteed that during this experiment your video recorded image will never be transmitted or

h w h a a ay' x .

PREDICTIVE MIND READING

U O U E R Y O Y U R E Y

*Display This Question:
If Treatment = None or Label*

Researchers have collected extensive data from the original experiment that was just described and are now interested to learn more about what you think participants did in the experiment. After being presented with a participant identification number and any relevant information about previous round choices and outcomes, you will have the opportunity to

“ Split” “ Take All”

correct guess can earn you \$1.00 and an incorrect guess can earn you \$0.00. You will have the opportunity to make 188 guesses across 3 parts of the survey. With a guess randomly chosen for payment from each part, you can earn a maximum of \$3.00 in addition to the wage you are guaranteed to earn by completing this survey.

*Display This Question:
If Treatment = Photo*

Researchers have collected extensive data from the original experiment that was just described and are now interested to learn more about what you think participants did in the experiment. We have prepared images (photos from videos) of the participants during each of their multiple rounds of interaction, taken moments after they stated their intentions, but

“ Split” “ Take All”

you (one at a time). Reminder: participants never saw any images or videos of each other. After being presented with a participant identification number, a photo of a participant from the original experiment, and any relevant information about previous round choices and outcomes, you will have the opportunity to guess whether the participant went on to choose “ Split” “ Take All”

\$1 00

guess can earn you \$0.00. You will have the opportunity to make 188 guesses across 3 parts of the survey. With a guess randomly chosen for payment from each part, you can earn a maximum of \$3.00 in addition to the wage that you are guaranteed to earn by completing this survey.

*Display This Question:
If Treatment = Video*

Researchers have collected extensive data from the original experiment that was just described and are now interested to learn more about what you think participants did in the experiment. We have prepared short videos of the participants during each of their multiple rounds of interaction, taken moments after they stated their intentions, but before making

“ Split” “ Take All”

time). Reminder: participants never saw any images or videos of each other. After being presented with a participant identification number, a video of a participant from the original experiment, and any relevant information about previous round choices and outcomes, you will have the opportunity to guess whether the participant w “ Split” “ Take All”

\$ 1 00

you \$0.00. You will have the opportunity to make 188 guesses across 3 parts of the survey. With a guess randomly chosen for payment from each part, you can earn a maximum of \$3.00 in addition to the wage that you are guaranteed to earn by completing this survey.

PREDICTIVE MIND READING

Survey Part 2: *first round* guesses

[This choice was looped over the 94 target players randomly]

of 94

Original Experiment Description

[When the mouse hovered over the bold text above, the text from *The First-Round Decision* block would appear as an overlay]

Display This Question:

If Treatment != Label

ID# is deciding what to do. What do you guess they will do in the first round?

Display This Question:

If Treatment = Label

ID# a <gender> is deciding what to do. What do you guess they will do in the first round?

Display This Question:

If Treatment = Photo

<insert photo>

Display This Question:

If Treatment = Video

<insert video – options below would not appear until the video ended>

Take All

Split

Congratulations, you have finished Part 2.

Next, you will be able to see what participants did in the *first-round* interactions and make guesses about what they do in their *second-round* interactions. A correct guess can earn you \$1.00. Ready to go to Part 3?

PREDICTIVE MIND READING

Survey Part 3: *second round guesses*

[This choice was looped over the 94 target players randomly]

of 94

Original Experiment Description

[When the mouse hovered over the bold text above, the text from The First-Round Decision block would appear as an overlay]

Display This Question:

If Treatment != Label

ID# is deciding what to do is deciding what to do.

They found out that in the *first* round their partner chose to **<split/take all>**.

In the first round Participant **ID#** chose to **<split/take all>**.

What do you guess they will do in the second round?

Display This Question:

If Treatment = Gender

ID #X, a **<gender>** is deciding what to do is deciding what to do.

They found out that in the *first* round their partner chose to **<split/take all>**.

In the first round Participant **ID#** chose to **<split/take all>**.

What do you guess they will do in the second round?

Display This Question:

If Treatment = Photo

<insert photo>

Display This Question:

If Treatment = Video

<insert video – options below would not appear until the video ended>

Take All

Split

Congratulations, you have finished Part 3.

In the next part of the survey, we ask you to answer a final set of questions, required for your completion of the survey. Ready to go to Part 4?

PREDICTIVE MIND READING

Survey Part 4

What is your age? [input number]

Which [,]

Before providing you feedback on your guess, we present to you a final set of statements concerning men and women and their relationships in contemporary society. Please indicate the degree to which you agree or disagree with each statement. Your answers will not affect your rating or payment.

- No matter how accomplished he is, a man is not truly complete as a person unless he has the love of a woman.
- Many women are actually seeking special favors, such as hiring policies that favor them over men, under the guise of asking for "equality."
- In a disaster, women ought not necessarily to be rescued before men.
- Most women interpret innocent remarks or acts as being sexist.
- Women are too easily offended.
- People are often truly happy in life without being romantically involved with a member of the other sex.
- Feminists are not seeking for women to have more power than men.
- Many women have a quality of purity that few men possess.
- Women should be cherished and protected by men.
- Most women fail to appreciate fully all that men do for them.
- Women seek to gain power by getting control over men.
- Every man ought to have a woman whom he adores.
- Men are complete without women.
- Women exaggerate problems they have at work.
- Once a woman gets a man to commit to her, she usually tries to put him on a tight leash.
- When women lose to men in a fair competition, they typically complain about being discriminated against.
- A good woman should be set on a pedestal by her man.
- There are actually very few women who get a kick out of teasing men by seeming sexually available and then refusing male advances.
- Women, compared to men, tend to have a superior moral sensibility.
- Men should be willing to sacrifice their own well-being in order to provide financially for the women in their lives.
- Women, as compared to men, tend to have a more refined sense of culture and good taste.

[All questions answered with 6-point Likert: disagree strongly, disagree somewhat, disagree slightly, agree slightly, agree somewhat, agree strongly]

PREDICTIVE MIND READING

Before participating in today's study, were you familiar with XXX University?

- No. I had no familiarity with XXX University.
- Yes, I was familiar with XXX University before, but I don't have personal connections to it.
- Yes, I was familiar with XXX University and have personal connections to it.

XXX U

- No. I had no familiarity with XXX University's students.
- Yes, I was familiar with XXX University's students before, but I don't have personal connections to them.
- Yes, I was familiar with XXX University and have personal connections to them.

Display This Question:

If Treatment = Photo

Or Treatment = Video

Have you ever seen any of the people shown in today's study before today?

- No. I did not recognize any of the people shown in the survey as people I have seen before today.
- Yes, I recognized a person or people shown in the survey as someone I have seen before today.

Payment Feedback

Based on the sum of your earnings from a randomly chosen guess in Part 1, a randomly chosen guess in Part 2, and a randomly chosen guess in Part 3 of the survey, you earned \$X in addition to your fixed payment for survey completion.