NBER WORKING PAPER SERIES

DO BELIEFS JUSTIFY ACTIONS OR DO ACTIONS JUSTIFY BELIEFS? AN EXPERIMENT ON STATED BELIEFS, REVEALED BELIEFS, AND SOCIAL-IMAGE MANIPULATION

James Andreoni
Alison Sanchez

Working Paper 20649
http://www.nber.org/papers/w20649

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2014

We thank Nageeb Ali, Richard Carson, Vincent Crawford, Matthew Goldman, Michael Kuhn, Erin Troland and seminar participants at the 2014 ESA Meetings, UC San Diego Economics Department, and UC San Diego Neural Interaction Lab for helpful discussions and comments. Paul Feldman and Vincent Leah-Martin provided excellent assistance in running the experiments. Sanchez acknowledges financial support from a National Science Foundation Graduate Research Fellowship (#2009081811), and Andreoni acknowledges financial support from the National Science Foundation (SES0962484 and SES1427355). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Do Beliefs Justify Actions or Do Actions Justify Beliefs? An Experiment on Stated Beliefs, Revealed Beliefs, and Social-Image Manipulation

James Andreoni and Alison Sanchez

NBER Working Paper No. 20649
October 2014
JEL No. C9,D03,D83

ABSTRACT

We study whether actions are justified by beliefs, as is usually assumed, or whether beliefs are justified by actions. In our experiment, subjects participate in a trust game, after which they have an opportunity to state their beliefs about their opponent's actions. Subsequently, subjects participate in a task designed to "reveal" their true beliefs. We find that subjects who make selfish choices and show strategic sophistication falsely state their beliefs in order to project a more favorable social image. By contrast, their "revealed" beliefs were significantly more accurate, which betrayed these subjects as knowing that their selfishness was not justifiable by their opponent's behavior.

James Andreoni
Department of Economics
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0508
and NBER
andreoni@ucsd.edu

Alison Sanchez
Dept of Economics 0508
9500 Gilman Drive
La Jolla, CA 92093
alsanche@ucsd.edu

"And if, to be sure, sometimes you need to conceal a fact with words, do it in such a way that it does not become known, or, if it does become known, that you have a ready and quick defence."

-Niccoló Machiavelli, 1522

We investigate if concerns of social image affect how individuals express their beliefs. It has already been established in several studies that individuals care about how others view them[1]. Our investigation centers on whether and how individuals mitigate selfish behavior by manipulating perceptions of their beliefs. We begin by demonstrating that individuals may not have a true preference for fairness or reciprocity, but rather are motivated by their concern over their social- or self- image. In other words, individuals have a preference for *appearing* to be fair, instead of actually behaving fairly. Further, this concern over image will not motivate selfish individuals to change their selfish behavior. Instead, image-conscious selfish individuals will maintain their selfish behavior, but go out of their way to create the false impression that they are cooperative. We then demonstrate the strategies by which image-conscious selfish individuals create the false impression that they are cooperative. We propose that individuals, with the intention of maintaining their social image, purposely manipulate information they know others might use to evaluate their actions. Specifically, we look at whether individuals who take a selfish action will subsequently lie about the beliefs they held when they took that action in a way that mitigates their selfish behavior.

[1] See for instance Andreoni and Bernheim, 2009; Bagwell and Bernheim, 1996; Glazar and Konrad, 1996; and Ireland, 1994.

The novelty of our paper is to compare two different belief elicitation methods to expose subjects' motivations and strategies. In our approach we look at the difference between *stated beliefs,* the beliefs subjects express when their social image is at stake, and *revealed beliefs*, the beliefs subjects express when their entire payoff is at stake. To our knowledge we are the first to introduce this method of comparing belief elicitation methods as well as introducing the revealed beliefs elicitation device itself. We conduct a modified trust game, after which we obtain the two measures of beliefs. First, we directly ask subjects to state the beliefs they held when they made certain decisions. We term these *stated beliefs*, as there are no monetary consequences for stating inaccurate beliefs. Second, we implement a belief elicitation device where subjects indirectly express their beliefs by placing bets on different game outcomes. We term these *revealed beliefs*, as subjects' entire payoff is as stake and subjects are unaware that their beliefs are being measured. We find that subjects who take a selfish action in the trust game subsequently exhibited large differences between their *stated* beliefs and their *revealed* beliefs. Selfish subjects stated that they believed their opponent also took a selfish action. However, in contrast with what they stated when asked directly, selfish subjects staked their payoff on the chance that their opponent took a cooperative action. Thus, these subjects were revealed to believe that their opponent was in fact cooperative. By comparison, subjects who took a cooperative action in the trust game do not display this difference between their stated and revealed beliefs. In order to rule out confusion as an explanation for the difference between stated and revealed beliefs we measure the decision-making sophistication of subjects. *Sophisticated* subjects are subjects who did not violate stochastic dominance in the revealed elicitation task[2]. Our key result is that we find that the subjects with the largest

[2] Please see Section II for a complete explanation of how subjects' sophistication was determined.

differences between their stated and revealed beliefs are the subjects who are both selfish and sophisticated.

We suggest that the reason for the disparity between the selfish subjects' stated and revealed beliefs is that these selfish individuals are precluded from using their action to signal a positive social image. Instead, the selfish players must use their stated beliefs to signal their motivation in making a selfish decision and in doing so, mitigate their selfish actions and preserve a positive social image. On the other hand, subjects who have taken the cooperative action can signal a positive social image with their action alone and thus, would have no need to alter their social image through their stated beliefs.

Underlying these behavioral propositions is the hypothesis that individuals have a skilled understanding of the way others think. In particular, we propose that individuals understand the cognitive processes fundamental in judgment and that it is this sophisticated understanding of how humans judge others that may explain why sophisticated-selfish subjects express two contradictory beliefs. We propose that the sophisticated-selfish subjects use stated beliefs as an opportunity to influence others' perception of their selfish decisions whereas the revealed beliefs capture their "true" beliefs. We find that first-movers believe that second-movers will be more generous when the second-movers believe the first-mover intended to be generous. A key finding is that sophisticated-selfish second-movers are able to anticipate this belief and are able to alter their stated beliefs to match this expectation.

Moreover, we do not find evidence in support of positive reciprocity. Even when players believe that their partner has been generous toward them, they do not positively reciprocate as would be predicted by many theories of reciprocity. Importantly, we find that these players go out of their way to state on the stated

belief elicitation that they believed the other player was unkind to them. Again, when we examine these same players' incentivized revealed beliefs, we find that they did in fact believe that their partner was generous and were willing to stake their entire payoff on this fact. Thus it seems unlikely that their selfish actions were driven by a preference for reciprocity. Rather, our results indicate that reciprocity may be followed only when there are no viable excuses to relieve the obligation to return kindness for kindness. When even under a small amount of social pressure, subjects will go out of their way to "blame the victim" to relieve themselves of the social responsibility to be generous.

Our results have important implications for policy. Previous research has focused on eliciting emotional reactions in individuals in order to influence charitable donations or encouraging adhesion to social norms in order to motivate prosocial behavior. In light of our results, the strategies promoted by previous work may, in fact, drive people *away* from other-regarding behavior.

Given that beliefs seem to be playing a larger role in theory and are being increasingly relied upon as an explanation for behavior, it seems prudent to examine whether beliefs are influenced by social demand. This demand can take the form of experimenter demand, audience demand, or demand of societal expectation. Perhaps subjects write beliefs that they want to use as "socially acceptable excuses" to validate and explain why they behaved as they did. Further, humans may know how to give the "right answers" to project their image of themselves to others (that is, manipulate others' perceptions. If this is indeed the case, careful attention must be paid to all of the incentives faced by individuals when measuring their beliefs.

The paper proceeds as follows: Section I provides background on social influence, social preferences and beliefs; Section II describes the experimental

design and predictions; Section III presents the results; Section IV provides discussion; and Section V concludes. Readers familiar with the literature on social preferences and beliefs may wish to skip directly to Section II.

I. Interaction of Social Preferences and Beliefs

Social pressure in varying forms, such as reciprocity and guilt aversion, has been shown to produce prosocial behavior in lab settings (Fehr and Schmidt, 1999; Charness and Rabin, 2002; Charness and Levin, 2007; Charness and Haruvey, 2002; Blount, 1993). However, previous research has taken for granted that individuals' only alternative in the face of social pressure is to behave cooperatively. Recent evidence has suggested that a concern for projecting a social image of being fair or reciprocal dominates concerns for reciprocity or fairness and those individuals only appear to have concerns regarding fairness, reciprocity and intentions. When confronted with an opportunity to give, a selfish person may give if they feel that their social or self image is at stake. These individuals may truly prefer to be self-regarding, but refrain under social pressure (Andreoni and Bernheim, 2009; Dana, et. al 2007). For example Andreoni and Bernheim (2009) use a variation of the Dictator game to test for social image concerns. In their design, nature sometimes intervenes on the Dictator's decision, forcing them to choose to keep the whole endowment. The recipient cannot observe if nature intervened. They found that when there was zero chance that any unfavorable outcome could be attributed to the external source (nature), Dictators ceded 50 percent of their endowment to the recipient. However, as soon as the chance of nature intervening reached 25 percent, Dictators' other-regarding behavior quickly reversed and many Dictators switched to choosing the entire endowment for themselves. The authors concluded that individuals have a strong preference for being perceived as being fair, which dominates their other-regarding concerns. In another relevant finding,

Dana, et. al. (2007) find that Dictators chose to remain ignorant of the consequences of their actions so as to retain the "moral wiggle room" to act selfishly.

In addition to preferences, beliefs have played an important role in explaining prosocial behavior[3]. Several past studies have incorporated beliefs about subject intentions in models of reciprocity (Fehr and Schmidt, 1999; Levine, 1998; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006). These models claim that individuals' actions are motivated by their beliefs about the intentions of others. Individuals will go out of their way to reward those who have intentionally shown them kindness and punish those who have intentionally hurt them even if doing so is costly to the rewarder/punisher (Falk and Fischbacher, 2000; Andreoni, Brown, and Vesterlund, 2002; Charness and Rabin, 2002; Charness and Levin, 2007; Charness and Haruvey, 2002; Blount, 1993). While previous studies of social preferences have made a point of using beliefs to provide motivation for other-regarding preferences, many of these papers have studied beliefs only indirectly or, if measured directly, only impose a small cost to subjects for lying. In a separate study, Dufwenberg et al. (2006) elicited

[3] Many theories of social preferences have allowed for belief-dependent motivation based off of Geanakopolos, Pearce and Stacchetti (1989) and Gilboa and Schmeidler (1988) who found that traditional methods were inadequate in representing preferences that exhibit belief-dependent motivations. Rabin's (1993) reciprocity theory, in which a Player's preferences over material payoff distributions are influenced by the co-players intentions, is a well-known application of "psychological" game theory. Several extensions of Rabin's 1993 theory, including Dufwenberg and Kirsteiger (2004) and Battigalli and Dufwenberg (2009) have illustrated the importance of incorporating updated beliefs, others' beliefs, and players' plans of how they intend to play. For example, Battigalli and Dufwenberg (2007) suggest that individuals care about what other people give to others in order to avoid aversive feelings of guilt based on co-players beliefs and expectations.

beliefs in one-shot public goods games to explore the impact of framing and to assess theories of reciprocity and guilt-aversion. They found that when players know they are expected to be other-regarding they give according to what they believe others expect of them.

II. Experimental Design and Predictions

Each session consisted of three stages. In the first stage, subjects play a modified Trust game with binary choices. We employ the strategy method: subjects were asked to make binding choices for different scenarios, and paid based on one randomly chosen scenario at the end of the session. All choices were made with paper and pencil. In the second and third stages we collect non-incentivized stated beliefs and incentivized elicited beliefs, respectively.

A. Stage 1: Choices

We use a variation of the Berg, Dickhaut, & McCabe Trust (1995) game that restricts all players to binary choices. This was done to facilitate the belief elicitation in Stage 3. In Stage 1 subjects are randomly divided into pairs and randomly assigned to roles as Player 1 (P1) or Player 2 (P2). To begin, \$10 is placed into player 1's "account". Player 1 now decides either to send the whole \$10 to player 2 or to send \$2 to player 2 and keep \$8. The amount sent to player 2 is tripled. Player 2 decides how much of the tripled transfer they received,x to return to player 1. Player 2 must decide between two options: (1) whether to return $x/2$ to player 1 and keep $x/2$, or, (2) return $x/6$ to player 1 and keep $5x/6$. Further, with probability $1-p$, player 1's choice determines the amount transferred to player 2 (either \$10 or \$2), and, with probability p nature intervenes and the "Experimenter" forces player 1 to send the whole \$10 to player 2. We examine choices for six different values of p, $p \in$ (0, 0.05, 0.20, 0.40, 0.60, and 1). The parameter p is common knowledge, but player 2 cannot

observe whether nature intervened. We employ the strategy method to elicit choices for all six values of p: each player makes choices on six different "Decision Sheets". Player 1 subjects make a total of 6 choices: one choice on each of 6 sheets by marking whether they would choose to send \$10 or \$2 to player 2 for each sheet (even the treatment where $p = 1$). By comparison, player 2 subjects make 12 total choices: a conditional choice for the possibility that \$10 is sent, and, a conditional choice for the possibility that \$2 is sent for each value of p. For ease of reference, a diagram of the game is provided in Figure 1.

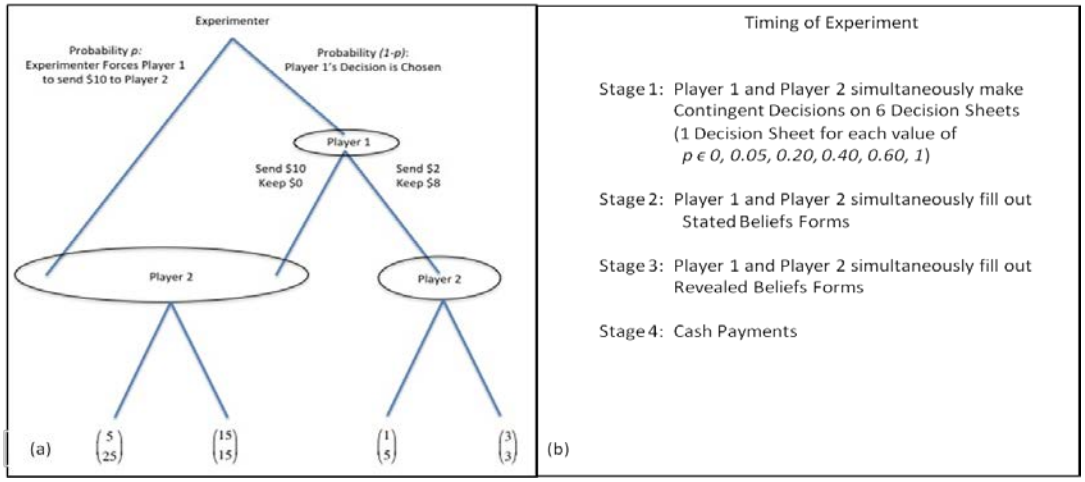

Figure 1 – (a) Modified Trust Game structure with payoffs. (b)Game Timing – All players move simultaneously without knowing what the other has chosen. Only after receiving payment in Stage 4 can players infer their opponent's moves.

B. Stage 2: Non-Incentivized "Stated" Beliefs

After all players made their choices, Decision Sheets were collected. Subjects were then told, "We would like to know what you think the other player sent you". The Belief Elicitation Stages (Stages 2 and 3) were not announced to the subjects until after each subject had made their choices in the first stage and the Decision Sheets were collected. Each subject wrote their predictions on their

own form called the Prediction Sheet. It was made clear to subjects that there was no penalty or reward for accuracy.

Player 1 subjects were asked to predict the chances that player 2 would send back different amounts of money to player 1 under three different scenarios. It was publically stated that (a) if $10 is sent to player 2, player 2 could send back either $15 or $5 to player 1, (b) if $2 is sent to player 2, player 2 could send back either $3 or $1 to player 1, and (c) player 2 makes a conditional choice for each possibility. It was also common knowledge that in each condition there existed a chance that nature could "override" player 1's choice and force player 1 to send $10 to player 2. Therefore, there were two beliefs player 2 could have held about who was responsible for sending the $10: first, player 1 was responsible themselves for sending the $10, or, second, player 1 was forced by the Experimenter to send the $10. This stage was designed to test if individuals operate on the assumption that their intentions will be taken into account when being judged by others. Therefore, player 1 was asked to predict the chances that player 2 sent back $15 or $5 for each of two possibilities: first, if player 2 believed player 1 was responsible for sending the $10 and second, if player 2 believed player 1 was not responsible for sending the $10 (i.e. player 1 was forced by the Experimenter). Player 1 was then asked to predict the chances that player 2 sent $3 or $1 back to player 1 under the $2 possibility. Player 1 made predictions for all six values of p. Again, it was made clear to subjects that there was no penalty or reward given for accuracy.

Player 2 subjects were asked to predict the chances that player 1 would send either $10 or $2. They were specifically instructed to predict the chances that player 1 chose $10 or $2 on their Decision Sheet and not the chances that player 2 would receive $10 or $2 (which depends on the chances that player 1's

decision is chosen). Player 2 made predictions for all six values of p. Again, there was no penalty or reward given for accuracy.

C. Stage 3: Incentivized "Revealed" Belief Elicitation

We employ a unique method in order to measure first-order beliefs about what player 1 and player 2 thought the other player had done. We use a Multiple Price List style approach to measure player 1 and player 2's preference between two payment options. An advantage of our method over other elicitation methods is that our method is not affected by risk aversion. Subjects were informed that they would be making a series of decisions on "how they would like to be paid" on a Payment Option Form. Option 1 is the "Outcome of the Game". If players choose this option, they are paid based on the outcome of the game they played with their opponent. Subjects knew that the payment they would receive under this option, either $\$x/6$ (the "low" amount) or $\$5x/6$ (the "high" amount), depended in part on what the other player chose to send them. Option 2 was a q chance of receiving $\$x/6$ and a $1-q$ chance of receiving $\$5x/6$. Option 2 varied in incremental steps of 5 percent, which ranged from a 0 percent chance of receiving $\$x/6$ and a 100 percent chance of receiving $\$5x/6$, to a 100 percent chance of receiving $\$x/6$ and a 0 percent chance of receiving $\$5x/6$. The two amounts of money, $\$x/6$ and $\$5x/6$, are the same two amounts of money the player could earn if they were to choose the Outcome of the Game option (Option 1). However, instead of their payment relying upon what the other player chose to send them (as under the Outcome of the Game Option), the payment under Option 2 depended solely upon the chances they saw listed under Option 2. Therefore, the row at which a subject decides to switch from Option 2 to Option 1 reveals the range of values of their belief about what the other player has chosen to send them. Subjects fill out one Payment Option Form for each of

their six Decision Sheets[4]. Since the first row under Option 2 gives the player a 100 percent chance of receiving the high amount, rational subjects who understand the game should initially prefer Option 2, if they believe that there is less than a 100 percent chance they will receive the high amount under Option 1.[5]

Whereas most previous studies infer beliefs solely from analyzing subjects' choices or use a scoring rule to elicit beliefs, we elicit beliefs not once, but twice, and exploit the difference between the two measures to expose subjects' motives. Our study differs from previous studies of beliefs which have implemented a scoring rule technique (quadratic loss function, etc.) to elicit accurate beliefs. Subjects in the studies utilizing a scoring rule receive a "bonus" from accurately reporting beliefs in addition to the money earned from playing any game in

[4] Previous price list style experiments have documented that a portion of subjects tend to switch multiple times between the two options presented (Holt and Laury, 2002; Meier and Sprenger, 2010; and Jacobsen and Petrie, 2009). It is generally accepted that since multiple switch points can indicate subject confusion and are difficult to rationalize, a framing device may be necessary to avoid subject confusion and clarify the decision process (Andreoni and Sprenger, 2012). We used animated instructions in order to illustrate the directions for the subjects. Out of 82 subjects, two subjects had multiple switch points on one or more of their Payment Option forms and one subject who switched "backwards" (starting with Option 1 and later switching to Option 2).

[5] Under Option 2, the probability of receiving the high amount declines with each descending row, while the probability of receiving the low amount increases with each descending row. At the row where a subject believes that they would have a higher probability of receiving the high amount from the other Player than the probability they see under Option 2, the subject has the incentive to switch to Option 1. Thus the row where each subject switches allows us to infer their belief about the chances of the other Player sending the high amount. In addition, we verbally instructed subjects that "Most people begin by preferring Option 2 and then switch to Option 1. Thus one way to view this task is to determine the best row to stop checking the box under Option 2 and start checking the box for Option 1"

which they had participated. Thus, only a small portion of each subject's payments come from accurately reporting beliefs. If any subject wished to conceal their true beliefs they would only need to sacrifice a small percentage of their entire payment to do so. So, while these scoring rules do offer monetary incentives to accurately report beliefs, there is no way to tell if subjects who wish to lie about their beliefs would give up their "bonus" in order to signal a false belief to the experimenter. In contrast, our method of eliciting revealed beliefs has higher stakes for the subjects. Each subject's entire payout for the experiment is determined from their decision on the revealed belief elicitation task. Furthermore, while it is obvious to subjects that researchers are collecting their beliefs when using a scoring rule, it is not obvious to subjects that we are measuring their beliefs on the revealed belief task. This is so for two reasons. First, subjects had just completed the stated belief task. Recall, the stated belief task was labeled as the "Prediction Sheet" on which the subjects were asked to make predictions about how the other player had behaved. On the Prediction Sheet, we asked subjects to tell us "What do you believe the other player sent to you?" Thus, the framing of the stated belief task made salient that we were inquiring about subjects' beliefs. In contrast, the aspect most salient for subjects in the revealed belief task was that their payment for the entire experiment was "on the line". Recall that the revealed belief task was labeled as the "Payment Option Form" on which subjects were instructed that "Now tell us how you would like to be paid." On the Payment Option Form, the subjects had two "payment" options. Subjects could either choose to be paid from the "Outcome of the Game" they had just played with their opponent, or, they could choose the outside gamble. Thus, the framing of the revealed belief task nudged subjects toward focusing on their payment.

After Stage 3, player 1 rolled the dice in order to determine whether it would be player 1's decision that would be chosen or if player 1's decision would be overridden (i.e., the Experimenter's decision would be used instead). In order to maintain anonymity, all subjects rolled the dice.

D. Predictions

Prediction 1- Social Image Perception Manipulation: *Individuals who exhibit selfish behavior in Stage 1 will justify this social behavior on their "stated" beliefs (Stage 2) by "blaming the victim."*

Prediction 2- Revealed Deceit: *Selfish individuals will exhibit a difference between what they "state" they believe on their stated beliefs and what they are "revealed" to believe on their revealed beliefs.*

We propose that selfish individuals deem that their selfish action will be evaluated in a kinder light if they are perceived as reacting to a belief that their opponent was selfish rather than if they are perceived as truly believing that their opponent acted kindly toward them and subsequently taking advantage of this kindness. Since selfish individuals can no longer use their actions to signal their type to the experimenter, they must rely on the only means left available to maintain their social image: others' perceptions of their beliefs. Consequently, selfish individuals wishing to maintain their social image will state on their "stated" beliefs that they believe that there is a low probability of their opponent voluntarily sending the high amount and a high probability that their opponent will send them the low amount. This serves as an excuse for a selfish individual's behavior. However, we predict that not all of these individuals truly believe that their opponents were selfish. We posit that selfish individuals will be willing to risk their entire payment for the experiment on their true belief that their opponent was kind to them by sending the high amount. Thus, revealing

that they believe they have a better chance of receiving the high amount from their opponent rather than from the outside gamble[6].

Prediction 3- Sophistication: *The selfish individuals who are more sophisticated will be the individuals most likely to lie about their stated beliefs.*

Our design allows us to separate out individuals who are more sophisticated decision makers. We term subjects who switched to Option 1 immediately on the last Payment Option form as being "Sophisticated"[7]. We predict that since these subjects are capable of understanding complex situations they are the subjects who would best be able to navigate situations in which they need to cover over their bad deeds. Examining the behavior of these "sophisticated" subjects can also help to rule out confusion as a factor in any differences seen between stated and revealed beliefs.

III. Main Findings

Eight-two subjects were recruited from the undergraduate population at University of California, San Diego. Each session was conducted at the Economics Laboratory at UC San Diego and ran between an hour and 30 minutes and two hours. Each subject maintained the same role (player 1 or player 2) throughout. Average earnings were $19, including a $7 participation fee, with a standard deviation of $8.16. Payoffs ranged from $10 to $32.

[6] Recall that the payment received under Option 1 is dependent on the action of their opponent. Therefore, the sooner a subject "switches" to Option 1 the higher is their belief about their opponent sending them the high amount.

[7] Recall that on the Stage 3 "Revealed" Belief elicitation form players are faced with two payment options: Option 1, receiving a payment from the outcome of the game played with their opponent and Option 2, receiving a payment from an outside gamble. Recall also that for the last Decision Sheet and corresponding Payment Option form the chance that the Experimenter will force player 1 to send the whole $10 to player 2 is 100 percent . Therefore, player 2 will receive $15 (the highest amount) with 100 percent probability. Therefore, it is in a player's best interest to switch to Option 1 (payment from the game) immediately since there is a 100 percent chance they will receive $15 from the game, while there is less than 100 percent chance they will receive $15 from the outside gamble.

A. Choice Behavior

Result 1a – Selfish Behavior: *As the chance that player 1 is forced by the Experimenter to send $10 increases, the fraction of player 2s returning $15 (an Equal split) declines steadily.*

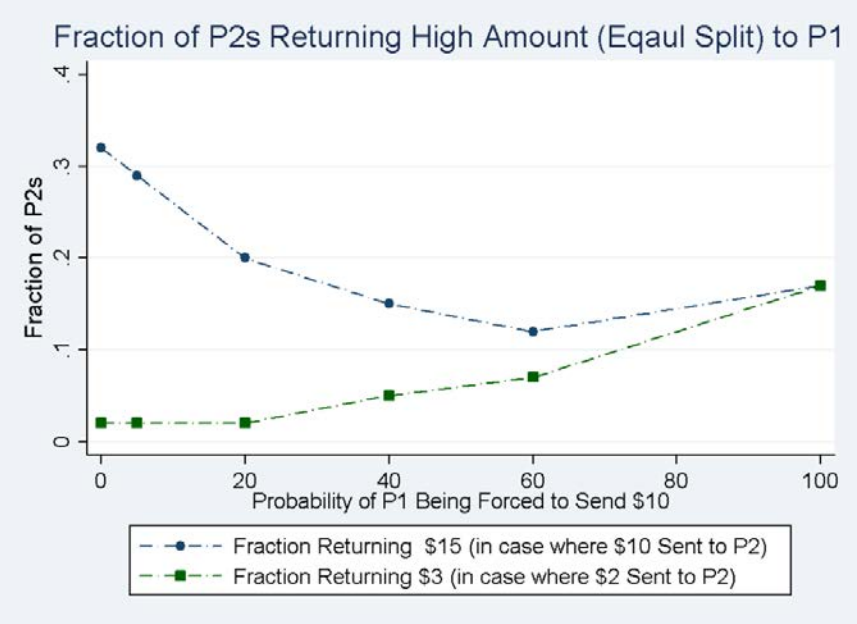

Figure 2 -Fraction of player 2s returning equal split to player 1 as the chance that player 1 was forced to be generous to player 2 increases

As can be seen in Figure 2 approximately 30 percent of player 2s return $15 to player 1 when (as was publically stated) there is zero chance that player 1 was forced to send $10. In this case ($p$ =0), player 2s know with certainty that if they receive $10 that it was player 1 who decided to send the $10 and it was of their own volition. Therefore, responsibility for sending the $10 is unambiguous to player 2s. However, as the chance that player 1 will be forced to send $10 increases, the fraction of player 2s reciprocating by returning an equal split of $15 declines. There is a small increase in the number of player 2s returning $15

on the last decision sheet, where the probability of player 1 being forced to send $10 reaches 100 percent.

The first column of Table 1 reports the estimates of a random-effects probit model of the probability of player 2 returning $15 in the case where $10 is sent to them. The second column reports the probability of player 2 returning $3 in the case where $2 is sent to them. The explanatory variables include indicators for $p \geq 0.05$, $p \geq 0.20$, $p \geq 0.40$, $p \geq 0.60$, and $p=1$ (with $p = 0$ omitted). In all cases, we report marginal effects at mean values. As we are most interested in player 2 reaction to knowledge that player 1 could have been forced to send $10, we focus on the results in the first column. The coefficients in the first column imply that there is a statistically significant decrease in the probability of player 2 returning $15 when p rises from 0.05 to 0.20, from 0.20 to 0.40, from 0.40 to 0.60 and from p=0.60 to 1.

Table 1

Probability of Player 2 Choosing Equal Split, Conditional on Probability of Player 1 Being Forced Random Effects Probit : Marginal Effects

Probability of Player 1 Being Forced to Send $10	(1) **If $10 Sent to Player 2:** Probability of Player 2 Returning **$15**	(2) **If $2 Sent to Player 2:** Probability of Player 2 Returning **$3**
p ≥ 0	-0.948**	-3.109***
	(0.417)	(0.855)
p ≥ 0.05	-0.137	0.000
	(0.393)	(0.751)
p ≥ 0.20	-0.717*	-0.210
	(0.422)	(0.817)
p ≥ 0.40	-1.105**	0.332
	(0.454)	(0.708)
p ≥ 0.60	-1.246***	0.665
	(0.453)	(0.674)
p= 1	-0.903**	1.341**
	(0.428)	(0.660)
Observations	246	246

Standard Errors in parentheses.
Significance*** $\alpha < 0.01$, ** $\alpha < 0.05$, * $\alpha < 0.10$

We now turn to player 1 choice behavior. Figure 3 shows the fraction of player 1s who voluntarily chose to send $10 to player 2. When the probability of being forced to send the whole $10 to player 2 is zero, around 30 percent of player 1s voluntarily choose to send $10. The fraction player 1s voluntarily choosing to send $10 increases gradually as the probability that they will be forced to do so increases. When the probability of being forced to send $10 is 100 percent, half of player 1s voluntarily choose to send $10. Table 2 shows the marginal effects from a random effects regression. The specification describes the probability of selecting $10. The explanatory variables include indicators for $p \geq 0.05$, $p \geq 0.20$, $p \geq 0.40$, $p \geq 0.60$, and $p=1$ (with $p = 0$ omitted). We report marginal

effects at mean values. The coefficients imply that the only statistically significant increase in the probability of voluntarily choosing to send $10 occurs when p rises from 0.60 to 1 ($\alpha < 0.10$, one tailed t-test).

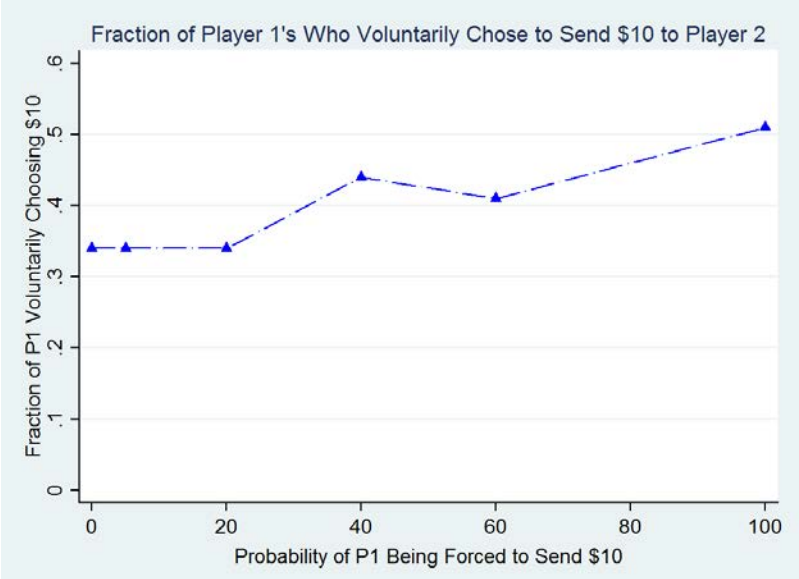

Figure 3- Fraction of player 1s who choose to be generous as probability that they will be forced increases

Table 2

Probability of Player 1 Voluntarily Sending $10
Random Effects Probit Model: Marginal Effects[a]

Probability of Player 1 Being **Forced** to Send $10	Probability of Player 1 **Voluntarily Choosing** to Send $10
$p \geq 0$	-0.368
	(0.273)
$p \geq 0.05$	-0.020
	(0.324)
$p \geq 0.20$	-0.113
	(0.327)
$p \geq 0.40$	-0.244
	(0.322)
$p \geq 0.60$	-0.154
	(0.322)
$p = 1$	0.538*
	(0.311)
Mean	0.167
Observations	246

[a]Standard Errors given in parentheses.
Significance: *** $\alpha < 0.01$, ** $\alpha < 0.05$, * $\alpha < 0.10$

B. Examing Beliefs: What Subjects <u>Say</u> They Believe Versus What They <u>Reveal</u> They Believe

Figure 4 shows stated beliefs, revealed beliefs and the actual frequency of player 1 sending $10 to player 2 (notice that this is the probability of player 1 voluntarily choosing $10 and not the probability that player 2 will receive $10). What is apparent from cursory examination is that there is a constant difference of approximately 20 percentage points between what player 2s *state* they believe

and what player 2s are revealed to believe. This difference is statistically significant for all six values of *p.*

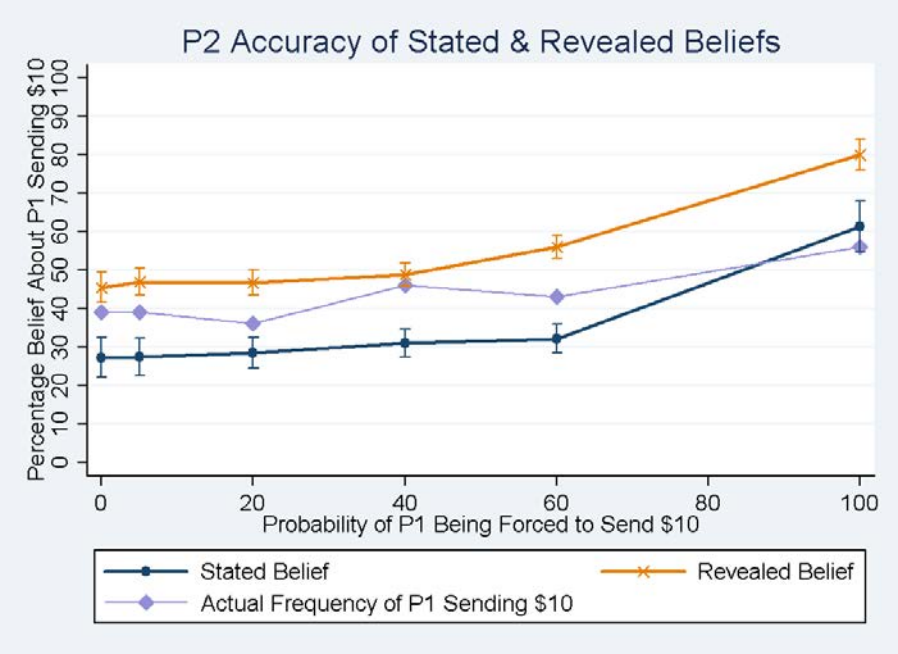

Figure 4 - Difference between player 2 "Stated" vs. true "Revealed" Beliefs

In order to further investigate the cause of this large difference between stated and revealed beliefs, we break the player 2s into four types based on 2 dimensions: Selfishness and Sophistication. We code a player 2 as being "Selfish" if they chose to send $5 (the lower amount) to player 1 for every value of *p*. Otherwise, the subject was coded as "Nice". We code a player 2 as being "Sophisticated" if on the Payment Option Form where $p = 1$, the subject

switched from preferring the Outside Gamble (Option 2) to preferring the Outcome of the Game (Option 1) in Row 1 or Row 2[8].

Table 3

Number Player 2-players in Each Category Type

	Unsophisticated	Sophisticated	Total
Nice	10 (24%)	8 (20%)	18 (44%)
Selfish	8 (20%)	15 (36%)	23 (56%)
Total	18 (44%)	23 (56%)	**41 (100%)**

Result 1b – Sophisticated Social Image Manipulation: *Sophisticated-Selfish player 2s are* revealed *to believe that there is a much higher chance that player 1 voluntarily sent $10 than they* ***state*** *they believe. Furthermore, the sophisticated-selfish player 2s are capable of accurately predicting the actual frequency that player 1 voluntarily chose $10, but when asked player 2s state a much lower probability than was true.*

Figure 5 shows the Stated Beliefs, Revealed Beliefs, and Actual Frequency of player 1 voluntarily choosing to send $10 for each type of player 2 (unsophisticated -nice, unsophisticated -selfish, sophisticated-nice, sophisticated-selfish). Comparing the Actual Frequency line with the Revealed Belief Line, one can see that sophisticated-selfish player 2s are fully capable of predicting player 1s' actions. In fact, there is no statistically significant difference between the Actual Frequency and the Revealed Belief for $p=0, p=0.05, p=0.20, p=0.40$. There is a significant difference for the last two values of p, $p=0.60$ and $p=1$.

[8] Subjects switching in Row 1 or Row 2 of the Payment Option form would have to be aware that on the last decision sheet, they were guaranteed to receive $10 as the probability of player 1 being forced to send $10 was 100 percent on this sheet.

There is a substantial increase in the Revealed Beliefs for the last two values of *p* which causes the difference. The sophisticated-selfish subjects do not best respond to their stated beliefs, but to their revealed beliefs. This indicates that the sophisticated-selfish subjects lie about their beliefs when asked, but do not believe their own lies. That is, they know what the Actual Frequency of player 1 voluntarily choosing $10 is, but appear to purposely understate this probability.

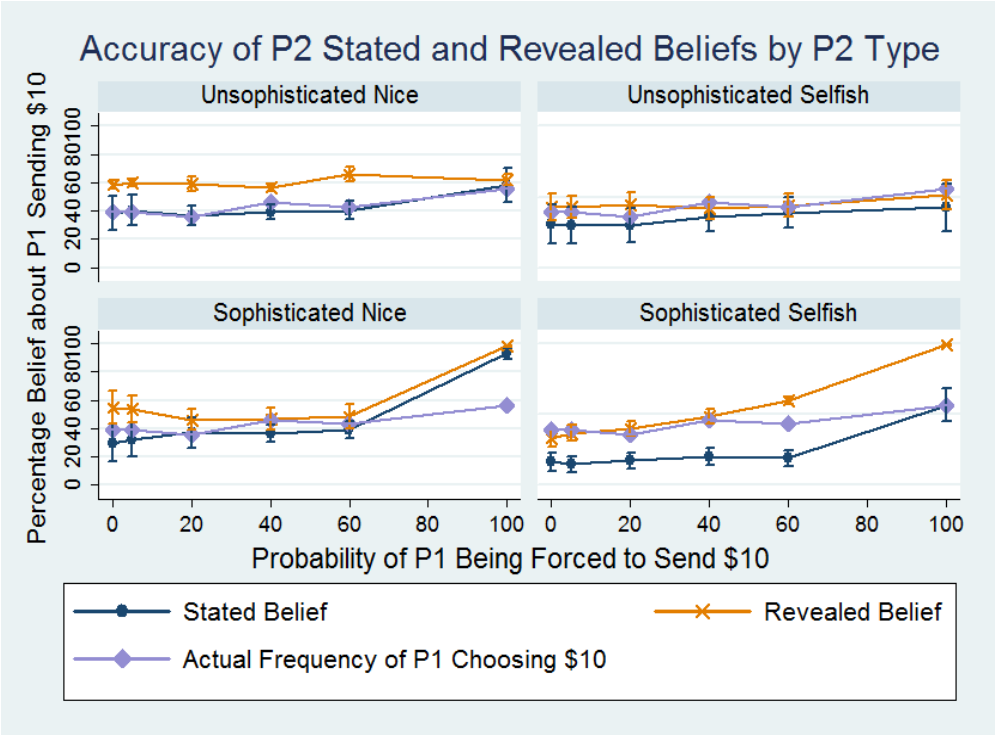

Figure 5 - Comparison of player 2 "Stated" vs. "Revealed" Beliefs broken down by Sophisticated & Selfish

Unsophisticated-nice subjects do exhibit a statistically significant difference for *p=0, p=0.05, p=0.20, p=0.40, p=0.60* and *p=1*. However, there is no statistically significant difference between unsophisticated-nice subjects' stated beliefs and the Actual Frequency. This leads one to conclude that the unsophisticated-nice subjects fail to best respond to their beliefs in a way that is overly optimistic. Sophisticated-nice subjects exhibit only occasional differences

between stated & revealed, revealed & actual frequency, and stated & actual frequency. On average, sophisticated-nice subjects best respond to their stated beliefs, but sometimes fail to best respond to their stated beliefs in favor of being optimistic. Unsophisticated-selfish subjects exhibit no statistically significant differences between stated & revealed beliefs, stated beliefs & actual frequency, and occasional significant differences between revealed beliefs & actual frequency. This indicates that unsophisticated-selfish subjects are both honest and realistic in that they truthfully state their beliefs and best respond to these beliefs.

As a further test, we compare player 2s who have large differences between their revealed and stated beliefs with player 2s who have little or no difference between what they say they believe and what they are revealed to believe. Those player 2s who are "large deviators" are significantly more selfish than those player 2s with small or no deviations ($t = 4.06, \alpha < 0.00$ two-tailed *t*-test, Mann-Whitney $z = 3.42, \alpha < 0.00$).

Now to shed further light on player 2 behavior, we contrast player 2 belief data with player 1 behavior on the two belief elicitation tasks. Looking at Figure 6, it can be seen that player 1s state that they believe that if player 2 believes player 1 is responsible for sending the $10 then player 2 will positively reciprocate. In Figure 7, we break Pl subjects into two groups, nice and selfish player 1s. player 1s who chose to send $10 to player 2 at least three times are coded as "Nice," otherwise they are coded as "Selfish." Figure 7 displays player 1 Stated Beliefs, Revealed Beliefs and the Actual Frequency of player 2 returning an equal split.

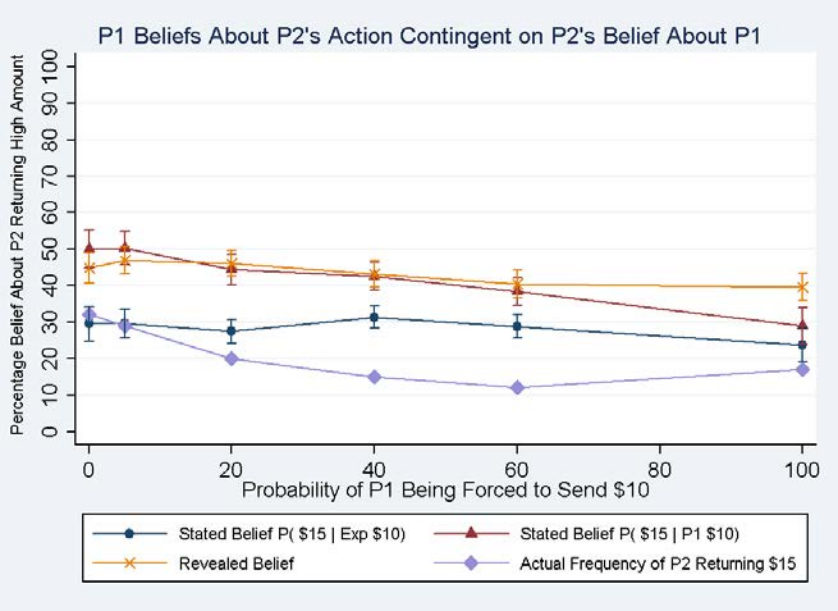

Figure 6 - Player 1 Beliefs about whether player 2 will return High Amount as the probability of player 1 being forced to be generous increases

Result 2- Sophisticated Deception: *Player 2s are able to accurately model player 1's expectations of player 2 behavior. Both types of player 1s both state and reveal that they believe that if player 2 believes that player 1 is responsible for voluntarily sending the $10, that player 2 will reciprocate this kindness by returning $15.*

Both types of player 1s state that they believe that $P(\ \$15\ |P1\ \$10) > P\ (\$15\ |Exp\ \$10)$ (pooled: $t = 5.79, \alpha < 0.00$, Selfish: $t = 4.90, \alpha < 0.00$, Nice: $t = 3.57, \alpha < 0.00$ two tailed *t*-tests). Now comparing Revealed Beliefs with Stated Beliefs one can see that not only are both types of player 1s truthful, but both types are operating on the assumption that player 2 will positively reciprocate if player 1 is perceived as responsible for voluntarily sending $10. There is no statistically significant difference between the Revealed Beliefs and

the Stated Belief of $P(\$15 \mid P1\ \$10)$ for both types of player 1s. This indicates that player 1s predict that player 2 will behave reciprocally, as is the social norm.

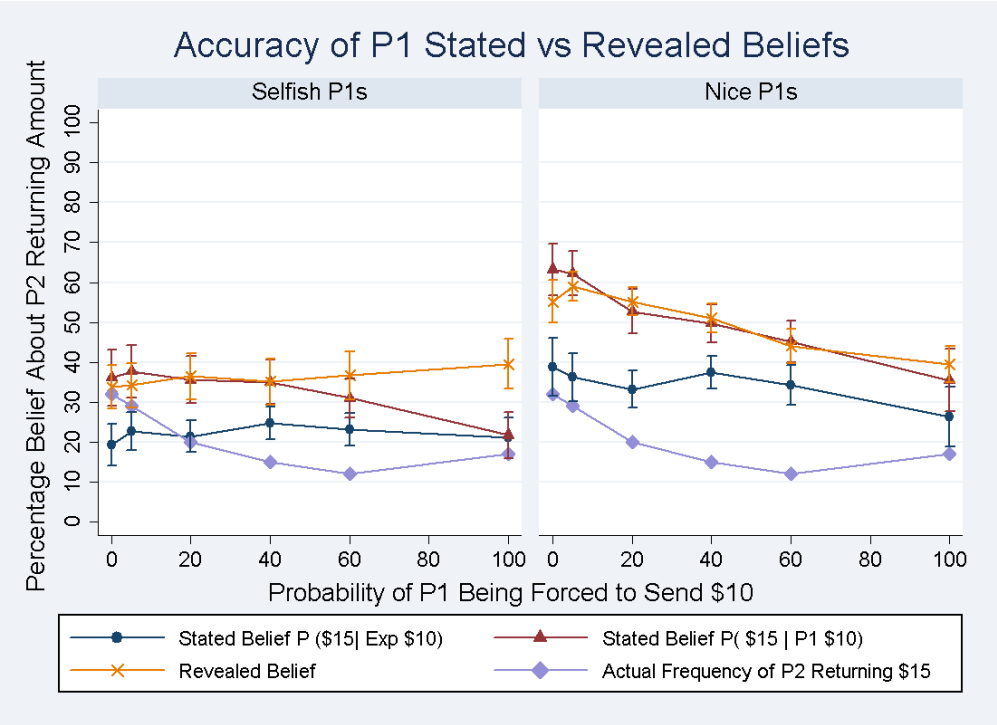

Figure 7 - Comparison of Selfish player 1s' and Nice player 1s' "Stated" vs. "Revealed" Beliefs

Result 3 : *Selfish player 1s understate their beliefs about* $P(\$15 \mid P1\ \$10)$.

The distribution of player 1s who had the largest difference between their revealed and stated beliefs are significantly more selfish than those player 1s who had little or no difference between their revealed and stated beliefs ($t = 1.96, \alpha < 0.025$ two tailed *t*-test). Looking at Figure 7, one can see that player 1s do not exhibit the same degree of deviation from their stated beliefs as exhibited by player 2s. Also evident from Figure 7 is that "nice" player 1s

believe there is a higher chance of receiving the high amount back from player 2 than do the selfish player 1s. Also, selfish player 1s were better at predicting the Actual Frequency that player 2 would return an equal split than are nice player 1s.

IV. Strategic Image Manipulation

The discrepancy between the beliefs selfish-sophisticated players express on their stated beliefs and their revealed beliefs raises several questions.

First, which belief represents players "true" beliefs? If sophisticated-selfish players' true beliefs are as they stated on their stated beliefs task, then these players are not best responding to these beliefs on their revealed beliefs task. If they truly believed that chances of player 1 sending the high amount ($10) were as low as they stated on their stated beliefs, then this should have been reflected by their choices on the Payment Option Form in the revealed beliefs task. Instead, their choices on their Payment Option Form indicate that their true underlying belief is that there was in fact a higher chance that player 1 had sent player 2 the high amount ($10). These sophisticated players are the players who were able to correctly calculate and assess that they were better off switching early from Option 2 to Option 1 on Decision Sheet 6, where the chance they would receive the high amount ($10) was 100 percent. Thus, it is difficult to comprehend how the sophisticated players can understand how to maneuver to make the most money for themselves in one part of the game and yet be confused or unable to best respond to their own beliefs on the very same task. In addition, selfish-sophisticated players demonstrated a keen ability to identify both the actual frequency of their opponents' kindness and their opponents' expectations. Those players who were selfish, but were unsophisticated, showed more consistency between their stated beliefs on the Prediction Sheet and their

revealed beliefs on their Payment Option Form. Lack of concern in social image may be possible explanation is for the difference in behavior between the sophisticated and unsophisticated -selfish players. It could be that unsophisticated players do not have enough knowledge to care about how others perceive them, or, they do care about how others perceive them but lack the prowess to manipulate others' perceptions.

The second question that arises is if their revealed beliefs are in fact more representative of their "true" beliefs, then what motivates sophisticated-selfish players to lie about their stated beliefs? We propose that sophisticated-selfish players intentionally misstate their stated beliefs in an effort to manipulate how others view their selfish actions. That is to say that sophisticated selfish subjects attempt to maintain a positive social image by manipulating how others perceive their selfish actions. Recall that on Figure 6 it was shown that player 1s stated that they believe that there was a higher chance of positive reciprocity from player 2 *if* player 2 believed that player 1, rather than the experimenter, was personally responsible for sending the high amount. Sophisticated-selfish stated beliefs coincide directly with their opponents' expectations. When directly asked about what they believed when calculating their decision of what amount to return to player 1, sophisticated-selfish player 2s stated that they believed that there was a low chance that player 1 had personally sent the high amount. As we know, this statement is in direct contradiction with their revealed beliefs. However, this statement does provide player 2 with a ready-made excuse should anyone inquire about their selfish decision[9]. The fact that sophisticated-selfish player 2s are so well able to anticipate their opponents' expectations exposes their keen awareness of and desire to appear to be in compliance with societal

[9] Concerns for social image maintenance may arise out of a desire to avoid social retaliation or revenge (see Andreoni and Gee (2012) for a review).

norms. If one were to look solely at sophisticated-selfish players' actions or their stated beliefs, it might appear that sophisticated-selfish players had preferences for reciprocity and that these preferences were their driving decision making process. However, the revealed beliefs paint a different picture entirely. Behavior that would previously had been viewed as supporting hypotheses of reciprocity or guilt-aversion is now shown to support the hypothesis that individuals are indeed selfish and display a sharp level of sophistication in manipulating their image. This is not to say that prosocial behavior in the form of pure or impure altruism does not exist. However, what this does imply is that if individuals wish to be selfish, attempts to nudge them to cooperate through appeals to reciprocity or guilt will not alter their choice behavior. In the face of social pressure to comply with norms sophisticated selfish individuals will not in fact cooperate, but merely take measures to make others believe they are complying with social norms.

The third question pertains to whether sophisticated-selfish players believe their own lies. We suggest that sophisticated-selfish players do not believe their own lies and that these players knowingly misstate the beliefs when they think that others will be able to observe their beliefs, as in the stated beliefs task. If sophisticated-selfish players truly believed that their opponent was selfish, then in order to have the best chance to earn the most money in the game they would have need to make entirely different choices on the Payment Option form in revealed belief elicitation task. Again, confusion seems unlikely as an explanation for their choices on this task since this set of players were the most sophisticated. In addition, their choices on the Payment Option Form in the revealed belief task match up with the actual frequency of cooperation from their opponents. This implies that sophisticated-selfish players are very apt at predicting exactly how kind their opponents were in the game. Therefore, it

seems unlikely that sophisticated-selfish players believed what they stated on their stated beliefs task. However, the question of whether the sophisticated players believe their own lies remains an open question and deserves more study.

A fourth and unanswered question is whether sophisticated-selfish subjects believe their own lies are credible and whether their lies are believed by observers? This remains an open question that deserves more study and is not directly addressed within our paper.

V. Conclusion

In order to examine whether image concerns affect how individuals express their beliefs we implement a new technique of contrasting two differing belief elicitation measures that identify both what people *say* they believe, and what they are revealed to believe. We find evidence that selfish, image-conscious individuals will lie about the beliefs they held about their opponent when carrying out this selfish action. Specifically, selfish players state that they believed their opponent would act selfishly toward them. However, in the revealed beliefs elicitation selfish players stake their entire payoff on the opposite belief, that their opponent was in fact cooperative. In order to rule out confusion, we measure the "sophistication" of each subject. Evidence indicates that it is the individuals who are both sophisticated and selfish who are the most frequent users of the manipulation mechanism. While previous studies of reciprocity concluded that players' beliefs about their opponents' intentions revealed that their subjects had a preference for reciprocity, our results contradict this finding. Our results suggest that individuals have a preference for being perceived as being cooperative instead of actually behaving cooperatively. Thus, if one wishes to take a selfish or uncooperative action, they can do so without

fear of retaliation or punishment so long as they concoct a socially acceptable story justifying their selfish behavior.

Our results have important policy implications. Previous studies have advocated appeals to individuals' emotions in order to promote cooperative behavior. However, our results indicate that guilt may in fact cause selfish individuals to not only act selfishly, but to cover up their selfish actions with lies. In fact, recent evidence from the healthcare field indicates that when doctors make their patients feel "guilty" about being overweight or other unhealthy behaviors a significant number of patients not only maintained their undesired behavior, but also lied and told their doctors that they had changed their behavior to the desired behavior (Darby, et.al. 2014).

We believe our results also serve as a caution for studying data in situations where social image is especially heightened. For example, the use of "big data" analytics from social networking websites has become popular. Large consumer firms and political campaigns have begun to rely on data gleaned from social media sites. However, since social image is particularly salient to individuals frequenting these websites, the truthfulness of their actions and statements on these websites may be suspect. Our results suggest that using data from these sources may lead to faulty conclusions given our evidence that people will create false impressions about themselves[10].

[10] For example, an industry of firms has emerged that helps individuals create false beliefs by employing "click farms" to create false Facebook "likes", to create false Twitter followers, or to create false LinkedIn "links". For a small fee a firm or individual can purchase Facebook "likes" or Twitter followers in order to create a false image of popularity or sphere of influence (VerSteeg and Galstayan, 2011; Wilbur and Zhu, 2009). Since social image is a powerful motivator, previous research has focused on using social media to influence behavior (encourage people to quit smoking, lose weight, exercise more frequently, vote a certain way, etc.). However in light of

Our results do not rule out cooperative behavior arising from altruism or warm-glow preferences. Since individuals can easily generate excuses to relieve themselves from social obligations to give, this supports evidence that giving is motivated by preferences for altruism and/or warm-glow. Further, people who want to behave selfishly will do so and that mechanisms designed to apply social pressure or guilt may do nothing to transform selfish behavior to cooperative behavior. Instead, selfish individuals may end up lying in order to maintain the *appearance* that they are cooperating with socially accepted group norms.

---

our findings, will individuals maintain their current "undesired" behavior and merely lie in order to maintain their social image?

References

Andreoni, J., P.M. Brown, L. Vesterlund. "What Makes an Allocation Fair ? Some Experimental Evidence." *Games and Economic Behavior* 40 (2002): 1-24.

Andreoni, James, and B. Douglas Bernheim. “Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects.” *Econometrica* 77, no. 5 (September 2009): 1607-1636.

Andreoni, James, and Charles Sprenger. “Uncertainty Equivalents: Linear Tests of the Independence Axiom.” *Mimeo, University of California, San Diego*, October 2010: 1-74.

Andreoni, James, and Laura Gee. "Gun For Hire: Delegated Enforcement and Peer Punishment in Public Goods Provision." *Journal of Public Economics* 96, no. 11-12 (December 2012): 1036–1046.

Andreoni, James, Justin M. Rao, and Hannah Trachtman. “Avoiding the Ask: A Field Experiment on Altruism, Emapthy and Charitable Giving.” *NBER Working Paper* 17648 (2011): 1-26.

Babcock, Linda, and George Lowenstein. “Explaining Bargaining Impasse: The Role of Self-Serving Bias.” *Journal of Economic Perspectives* 11, no. 1 (Winter 1997): 109-126.

Battigalli, P, and M. Dufwenberg. “Dynamic psychological games.” *Journal of Economic Theory*, no. 144 (2009): 1-35.

Bénabou, R., and J. Tirole. "Incentives and Prosocial Behavior." *American Economic Review* 96, no. 5 (2006): 1652--1678.

Bolton, Gary E., and Axel Ockenfels. “ERC: A Theory of Equity Reciprocity and Competition.” *American Economic Review* 100 (March 2000): 166–193.

Camerer, Colin F. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton, New Jersey: Princeton Univeristy Press, 2003.

Charness, Gary, and David I Levine. “Intention and Stocahstic Outcomes: An Experimental Study.” *The Economic Journal* 117, no. July (July 2007): 1051-1072.

Charness, Gary, and Ernan Haruvy. “Altruism, equity, and reciprocity in a gift-exchange experiment:an encompassing approach.” *Games and Economic Behavior* 40 (2002): 203-231.

Charness, Gary, and Matthew Rabin. “Understanding Social Preferences with Simple Tests.” *Quarterly Journal of Economics* 117, no. 3 (August 2002): 817-869.

Costa-Gomes, M, and V.P. Crawford. “Cognition and Behavior in Two-Person Guessing Games: An Experimental Study.” *The American Economic Review* 96, no. 5 (December 2006): 1737-1768.

Costa-Gomes, Miguel A., and Georg Weizsacker. “Stated Beliefs and Play in Normal-Form Games.” *Review of Economic Studies* 75 (2008): 729-762.

Crawford, V.P, M. Costa-Gomes, and N. Iriberri. “Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications.” *Journal of Economic Literature* LI, no. 1 (March 2013): 5-62.

Dana, Jason, Jason Xi Kuang, and Roberto A Weber. “Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness.” *Economic Theory* 33 (2007): 67-80.

Darby, R.S., N Henniger & C. Harris. “Shame and Guilt: Differential Health Outcomes in Physician-Patient Interactions.” *Basic and Applied Social Psychology*, In Press.

Dufwenberg, Martin, and Uri Gneezy. “Measuring Beliefs in an Experimental Lost Wallet Game.” *Games and Economic Behavior* 30 (2000): 163-182.

Erev, Ido, Gary Bornstein, and Thomas S. Wallsten. “The Negative Effect of Probability Assessments on Decision Quality.” *Organizational Behavior and Human Decision Processes* 55 (1993): 78-94.

Falk, Armin, and Urs Fischbacher. “A Theory of Reciprocity.” *Games and Economic Behavior* 54 (2006): 293-315.

Fehr, Ernst, and Klaus M. Schmidt. “A Theory of Fairness, Competition and Cooperation.” *Quarterly Journal of Economics* 114 (1999): 817-868.

Gachter, Simon, and Elke Renner. “The effects of (incentivized) belief elicitation in public goods experiments.” *Experimental Economics* 13 (2010): 364-377.

Geneseove, David, and Mayer Chris. “Loss Aversion and Seller Behavior: Evidence From the Housing Market.” *Quarterly Journal of Economics*, 2001, 116(4): 1233-60.

Johnson, N, and A Mislin. “Trust Games: A meta-analysis.” *Journal of Economic Psychology*, no. 32 (2011): 865-889.

McCabe, K, M Rigdon, and V.L. Smith. “Positive Reciprocity and intentions in trust games.” *Journal of Economic Behavior and Organization* 52 (2003): 267-275.

Rabin, Matthew. “Incorporating Fairness into Game Theory and Economics.” *American Economic Review* 83, no. 5 (December 1993): 1281-1302.

Rabin, Matthew. “Incorporating Fairness Into Game Theory and Economics.” *American Economic Review*, 1993: 83, 1281-1302.

Sobel, Joel. “Interdependent Preferences and Reciprocity.” *Journal of Economic Literature* 43 (June 2005): 392-436.

Tadelis, Steven. “The Power of Shame and Rationality of Trust.” *Working Paper, Haas School of Business UC Berkeley*, March 2011: 1-24.

VerSteeg, G. and A. Galstayan. “Information Transfer in Social Media.” *arXiv:1110.2724*, October 2011.

Zhu, Wilbur. K and Y. “Click Fraud.” *Marketing Science* 28, no. 2 (March-April 2009): 293-308.

Appendix A: Tables

Table 4 – Fixed Effects Regressions

	Player 1 Revealed Beliefs					
	(1) P1s Who Chose $10: Revealed Belief	(2) P1s Who Chose $2: Revealed Belief	(3) All P1s: Revealed Belief	(4) All P1s: Revealed Belief	(5) All P1s: Revealed Belief	(6) All P1s: Revealed Belief
Stated Belief About P($15| Exp Sent $10)	0.175***	0.141	0.110	0.112		0.138
	(0.064)	(0.137)	(0.117)	(0.114)		(0.090)
Stated Belief About P ($15 | P1 Sent $10)	0.330***	-0.170	0.125**	0.102*		0.419**
	(0.075)	(0.108)	(0.062)	(0.060)		(0.210)
Stated Belief About P($3| P1 Sent $2)				-0.143**		-0.695***
				(0.063)		(0.232)
Stated Belief About P($5| Exp Sent $10)					-0.064	0.043
					(0.112)	(0.066)
Stated Belief About P($5| P1 Sent $10)					-0.160***	0.302
					(0.059)	(0.209)
Stated Belief About P($1| P1 Sent $2)					0.074	-0.557***
					(0.072)	(0.207)
Constant	31.173***	38.362***	35.268***	38.227***	51.057***	59.068***
	(5.389)	(2.244)	(3.689)	(3.753)	(11.221)	(6.567)
Observations	97	135	232	232	232	232
Clusters	33	36	39	39	39	39
R-Squared	0.278	0.061	0.187	0.060	0.098	0.051

Robust Standard Errors in parentheses
Significance: *** $\alpha < 0.01$, ** $\alpha < 0.05$, * $\alpha < 0.10$

Player 2 Revealed vs Stated Beliefs For Each Value of *P=p_0 Player 1 Forced to Send $10*

p=0	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Revealed Belief	58.5	42.8125	54.688	31.167
Stated Belief	38.5	30.625	30	16.333
Difference	20*	12.1875	24.688*	14.834**
p=0.05	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Revealed Belief	60	43.125	53.75	36.833
Stated Belief	40.5	30	32.125	14.6
Difference	19.5**	13.125	21.625*	22.233***
p=0.20	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Revealed Belief	59.25	44.375	46.25	40
Stated Belief	36.8	30	37.125	17.333
Difference	22.45***	14.375	9.125	22.667***
p=0.40	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Revealed Belief	56.5	41.875	46.875	48.214
Stated Belief	39.5	35.625	36.625	20
Difference	17***	6.25	10.25	28.214***
p=0.60	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Revealed Belief	65.75	43.75	48.75	59.833
Stated Belief	40.5	38.75	39.375	19
Difference	25.25***	5	9.375	40.833***
p=1	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Revealed Belief	62	51.25	98.438	99.107
Stated Belief	58	42.5	93	56.733
Difference	4	8.75	5.438*	42.374***

Significance: *** $\alpha < 0.01$, ** $\alpha < 0.05$, * $\alpha < 0.10$

Player 2 Revealed vs True Prob Player 1 Voluntarily Chooses $10 For Each Value of *P=$p_0$ Player 1 Forced to Send $10*

p=0	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Revealed Belief	58.5	42.8125	54.688	31.167
True Prob Player 1 Chooses $10	39	39	39	39
Difference	19.5***	3.8125	15.688*	-7.833

p=0.05	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Revealed Belief	60	43.125	53.75	36.833
True Prob Player 1 Chooses $10	39	39	39	39
Difference	21***	4.125	14.75*	-2.167

p=0.20	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Revealed Belief	59.25	44.375	46.25	40
True Prob Player 1 Chooses $10	36	36	36	36
Difference	23.25***	8.375	10.25*	4

p=0.40	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Revealed Belief	56.5	41.875	46.875	48.214
True Prob Player 1 Chooses $10	46	46	46	46
Difference	10.5***	-4.125	0.875	2.214

p=0.60	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Revealed Belief	65.75	43.75	48.75	59.833
True Prob Player 1 Chooses $10	43	43	43	43
Difference	22.75***	0.75	5.75	16.833***

p=1	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Revealed Belief	62	51.25	98.438	99.107
True Prob Player 1 Chooses $10	56	56	56	56
Difference	6*	-4.75	42.438***	43.107***

Significance: *** $\alpha < 0.01$, ** $\alpha < 0.05$, * $\alpha < 0.10$

Player 2 Stated Belief vs True Prob Player 1 Voluntarily Chooses $10 For Each Value of *P=$p_0$ Player 1 Forced to Send $10*

p=0	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Stated Belief	38.5	30.625	30	16.333
True Prob Player 1 Chooses $10	39	39	39	39
Difference	-0.5	-8.375	-9	-22.667***

p=0.05	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Stated Belief	40.5	30	32.125	14.6
True Prob Player 1 Chooses $10	39	39	39	39
Difference	1.5	-9	-6.875	-24.4***

p=0.20	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Stated Belief	36.8	30	37.125	17.333
True Prob Player 1 Chooses $10	39	39	39	39
Difference	-2.2	-9	-1.875	-21.667***

p=0.40	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Stated Belief	39.5	35.625	36.625	20
True Prob Player 1 Chooses $10	39	39	39	39
Difference	0.5	-3.375	-2.375	-19***

p=0.60	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Stated Belief	40.5	38.75	39.375	19
True Prob Player 1 Chooses $10	39	39	39	39
Difference	1.5	-0.25	0.375*	-20***

p=1	**Unsophisticated Nice**	**Unsophisticated Selfish**	**Sophisticated Nice**	**Sophisticated Selfish**
Stated Belief	58	42.5	93	56.733
True Prob Player 1 Chooses $10	39	39	39	39
Difference	19	3.5	54***	17.733

Significance: *** $\alpha < 0.01$, ** $\alpha < 0.05$, * $\alpha < 0.10$

Player 1 Stated Belief About P ($15 | Player 1 $10) vs Revealed Belief

	Player 1s Who Chose $2	Player 1s Who Chose $10
p= 0		
Stated Belief P ($15 | Player 1 $10)	33.5	65
Revealed	37.28	72.857
Difference	-3.78	-7.857
p= 0.05		
Stated Belief P ($15 | Player 1 $10)	37.8	63.214
Revealed	37.6	72.857
Difference	0.2	-9.643*
p= 0.20		
Stated Belief P ($15 | Player 1 $10)	39.3	58.214
Revealed	36.24	58.929
Difference	3.06	-0.715
p= 0.40		
Stated Belief P ($15 | Player 1 $10)	32.5	55.833
Revealed	33.33	53.333
Difference	-0.83	2.5
p= 0.60		
Stated Belief P ($15 | Player 1 $10)	35.227	47.2
Revealed	35.863	41.412
Difference	-0.636	5.788
p= 1		
Stated Belief P ($15 | Player 1 $10)	35.8	42.738
Revealed	25.94	31.429
Difference	9.86	11.309*

Significance: *** $\alpha < 0.01$, ** $\alpha < 0.05$, * $\alpha < 0.10$

Player 1 Stated Belief About P ($15 | Player 1 $10) vs P ($15 | Exp $10)

	Player 1s Who Chose $2	Player 1s Who Chose $10
p= 0		
Stated Belief P ($15 | Player 1 $10)	37.28	72.857
Stated Belief P ($15 | Exp $10)	25.84	36.538
Difference	11.44*	36.319***
p= 0.05		
Stated Belief P ($15 | Player 1 $10)	37.64	72.857
Stated Belief P ($15 | Exp $10)	30.24	28.571
Difference	7.4	44.286***
p= 0.20		
Stated Belief P ($15 | Player 1 $10)	36.24	58.929
Stated Belief P ($15 | Exp $10)	25.24	31.429
Difference	11**	27.5***
p= 0.40		
Stated Belief P ($15 | Player 1 $10)	33.333	53.333
Stated Belief P ($15 | Exp $10)	24.524	39.167
Difference	8.809*	14.166**
p= 0.60		
Stated Belief P ($15 | Player 1 $10)	35.863	41.412
Stated Belief P ($15 | Exp $10)	23.818	35.294
Difference	12.045**	6.118
p= 1		
Stated Belief P ($15 | Player 1 $10)	25.941	31.429
Stated Belief P ($15 | Exp $10)	21.389	25.714
Difference	4.552	5.715

Significance: *** $\alpha < 0.01$, ** $\alpha < 0.05$, * $\alpha < 0.10$

Player 1 Stated Belief About P ($15 | Player 1 $10)
vs.
Actual Frequency of Player 2 Returning $15

	Player 1s Who Chose $2	Player 1s Who Chose $10
p= 0		
Actual Frequency of Player 2 Returning $15	32	32
Stated Belief P ($15 | Exp $10)	25.84	36.538
Difference	6.16*	-4.538
p= 0.05		
Actual Frequency of Player 2 Returning $15	29	29
Stated Belief P ($15 | Exp $10)	30.24	28.571
Difference	-1.24	0.429
p= 0.20		
Actual Frequency of Player 2 Returning $15	20	20
Stated Belief P ($15 | Exp $10)	25.24	31.429
Difference	-5.24*	-11.429**
p= 0.40		
Actual Frequency of Player 2 Returning $15	15	15
Stated Belief P ($15 | Exp $10)	24.524	39.167
Difference	-9.524***	-24.167***
p= 0.60		
Actual Frequency of Player 2 Returning $15	12	12
Stated Belief P ($15 | Exp $10)	23.818	35.294
Difference	-11.818***	-23.294***
p= 1		
Actual Frequency of Player 2 Returning $15	17	17
Stated Belief P ($15 | Exp $10)	21.389	25.714
Difference	-4.389	-8.714*

Significance: *** $\alpha < 0.01$, ** $\alpha < 0.05$, * $\alpha < 0.10$

Appendix B: Subject Forms

PLAYER 1 Decision Sheet 1 Subject # _____

PLAYER 1: Mark your decision in the right column labeled PLAYER 1's DECISION, where you can choose to send EITHER $10 or $2. (The Experimenter's Decision has already been "Marked").

EXPERIMENTER'S DECISION	PLAYER 1's DECISION
The Odds are: 0-in-100 (0 %) that Experimenter's Decision will be chosen	**The Odds are: 100-in-100 (100 %)** that Player 1's Decision will be chosen
The Experimenter always sends $10. **A. ☑ The Experimenter has required Player 1 to send $ 10** to Player 2, and Player 1 keeps **$ 0** . The **$10** that is sent will be TRIPLED, so Player 2 will actually get **$ 30** for their Account. From that **$ 30**, Player 2 can decide to do **one** of the following: 1) Player 2 can send **$15** back to Player 1 and keep **$15** for Themselves ... Or 2) Player 2 can send **$5** back to Player 1 and keep **$25** for Themselves.	**Player 1: Mark either $10 or $2** **A. ☐ I Choose to Send $10** to Player 2 and Keep **$ 0** for myself. The **$10** that I send will be TRIPLED, so Player 2 will actually get **$30** for their Account. From that **$30**, Player 2 can decide to do one of following: 1) Player 2 can send **$15** back to Player 1 and keep **$15** for Themselves ...Or 2) Player 2 can send **$5** back to Player 1 and keep **$25** for Themselves. **OR** **B. ☐ I Choose to Send $ 2** to Player 2 and Keep **$ 8** for myself. The **$2** that I send will be TRIPLED, so Player 2 will actually get **$ 6** for their Account. From that **$6**, Player 2 can decide to do one of the following: 1) Player 2 can send **$ 3** back to Player 1 and keep **$ 3** for Themselves, ...Or 2) Player 2 can send **$ 1** back to Player 1 and keep **$ 5** for Themselves

Player 2's Response *to* Decision Sheet 1

Subject #_____

You make a Choice for **BOTH** Decision A **and** Decision B: Choose "how much to send back to Player 1".

Decision A

represents the possibility that $10 is the amount sent to you on this Decision Sheet .

If $**10** is the amount sent , then the $**10** will be *Tripled*, and you will get $**30** for your Account. But, you won't know if it came from the **Experimenter's** Decision OR from **Player 1**'s Decision. However, **the Odds** are below (showing you the "chances" of Whose Decision is chosen):

the Odds are: 0-in-100 (**0 %**) the **Experimenter's Decision** is chosen

the Odds are: 100-in-100 (**100 %**) that **Player 1's Decision** is chosen

Player 2's Response to Decision A:

From the $ 30 ... I will Decide to send *Back to Player 1* (choose one of the following):

☐ Send $ **15** back to Player 1 and Keep $ **15** for Myself

OR

☐ Send $ **5** back to Player 1 and Keep $ **25** for Myself

Decision B

represents the possibility that $ 2 is the amount sent to you on this Decision Sheet .

If $**2** is the amount sent, then the $**2** will be *Tripled*, and you will get $ **6** for your Account. Remember, if $**2** is the amount sent to you, then you will know that **Player 1's Decision** is chosen, because **only Player 1** can send you $**2**.

Player 2's Response to Decision B:

From the $ 6 ... I will Decide to send *Back to Player 1* (choose one of the following):

☐ Send $ **3** back to Player 1 and Keep $ **3** for Myself

OR

☐ Send $ **1** back to Player 1 and Keep $ **5** for Myself

PLAYER 1 **PREDICTION SHEET** **SUBJECT_____**

Here are the ODDS
that Player 2 saw when deciding How Much to send back to you...

	If $10 is sent to Player 2, *and* Player 2 believes that it came from...				If $2 is sent to Player 2	
	Experimenter's Decision then I predict :		Player 1's Decision then I predict :		then I predict :	
	the Chances that Player 2 Sent $15 Back to Me are....	the Chances that Player 2 Sent $5 Back to Me are....	the Chances that Player 2 Sent $15 Back to Me are....	the Chances that Player 2 Sent $5 Back to Me are....	the Chances that Player 2 Sent $3 Back to Me are....	the Chances that Player 2 Sent $1 Back to Me are....
Decision Sheet 1 Odds for whose decision it is 0 % Experimenter's Decision 100 % Player 1's Decision	%	%	%	%	%	%
Decision Sheet 2 Odds for whose decision it is 5 % Experimenter's Decision 95 % Player 1's Decision	%	%	%	%	%	%
Decision Sheet 3 Odds for whose decision it is 20 % Experimenter's Decision 80 % Player 1's Decision	%	%	%	%	%	%
Decision Sheet 4 Odds for whose decision it is 40 % Experimenter's Decision 60 % Player 1's Decision	%	%	%	%	%	%
Decision Sheet 5 Odds for whose decision it is 60 % Experimenter's Decision 40 % Player 1's Decision	%	%	%	%	%	%
Decision Sheet 6 Odds for whose decision it is 100 % Experimenter's Decision 0 % Player 1's Decision	%	%	%	%	%	%

Player 2 **PREDICTION SHEET** **SUBJECT _____**

IMPORTANT

On each decision sheet, the Experimenter made a decision and Player 1 made a decision. BUT HERE, we want you to think ONLY about PLAYER 1's decision. On each decision sheet, Player 1 could have decided to send either $10 or $2 to you. What do you think the chances are that PLAYER 1 decided to send to you $10 as compared to $2 ?

Here are the ODDS that Player 1 saw when deciding How Much to send you... ↓	I Predict the chances that Player 1 chose to send $10 to me are...	I Predict the chances that Player 1 chose to send $2 to me are...
Decision Sheet 1 Odds for whose decision it is : 0% Experimenter's Decision 100 % Player 1's Decision	%	%
Decision Sheet 2 Odds for whose decision it is : 5% Experimenter's Decision 95 % Player 1's Decision	%	%
Decision Sheet 3 Odds for whose decision it is : 20% Experimenter's Decision 80% Player 1's Decision	%	%
Decision Sheet 4 Odds for whose decision it is : 40% Experimenter's Decision 60% Player 1's Decision	%	%
Decision Sheet 5 Odds for whose decision it is : 60% Experimenter's Decision 40% Player 1's Decision	%	%
Decision Sheet 6 Odds for whose decision it is : 100% Experimenter's Decision 0% Player 1's Decision	%	%

Player 2 | Payment Option Form | Subject _____

Option 1			Or	Option 2		
I want to go with the **Outcome of the Game** where I'll get either $5 or $15 depending on what **Player 1** sent me				I would rather go with these Odds here of ... **Chance of $5** ↓	**Chance of $15** ↓	
1)	*Outcome of Game*	☐	*Or*	**0 *in* 100**	**100 *in* 100**	☐
2)	*Outcome of Game*	☐	*Or*	**5 *in* 100**	**95 *in* 100**	☐
3)	*Outcome of Game*	☐	*Or*	**10 *in* 100**	**90 *in* 100**	☐
4)	*Outcome of Game*	☐	*Or*	**15 *in* 100**	**85 *in* 100**	☐
5)	*Outcome of Game*	☐	*Or*	**20 *in* 100**	**80 *in* 100**	☐
6)	*Outcome of Game*	☐	*Or*	**25 *in* 100**	**75 *in* 100**	☐
7)	*Outcome of Game*	☐	*Or*	**30 *in* 100**	**70 *in* 100**	☐
8)	*Outcome of Game*	☐	*Or*	**35 *in* 100**	**65 *in* 100**	☐
9)	*Outcome of Game*	☐	*Or*	**40 *in* 100**	**60 *in* 100**	☐
10)	*Outcome of Game*	☐	*Or*	**45 *in* 100**	**55 *in* 100**	☐
11)	*Outcome of Game*	☐	*Or*	**50 *in* 100**	**50 *in* 100**	☐
12)	*Outcome of Game*	☐	*Or*	**55 *in* 100**	**45 *in* 100**	☐
13)	*Outcome of Game*	☐	*Or*	**60 *in* 100**	**40 *in* 100**	☐
14)	*Outcome of Game*	☐	*Or*	**65 *in* 100**	**35 *in* 100**	☐
15)	*Outcome of Game*	☐	*Or*	**70 *in* 100**	**30 *in* 100**	☐
16)	*Outcome of Game*	☐	*Or*	**75 *in* 100**	**25 *in* 100**	☐
17)	*Outcome of Game*	☐	*Or*	**80 *in* 100**	**20 *in* 100**	☐
18)	*Outcome of Game*	☐	*Or*	**85 *in* 100**	**15 *in* 100**	☐
19)	*Outcome of Game*	☐	*Or*	**90 *in* 100**	**10 *in* 100**	☐
20)	*Outcome of Game*	☐	*Or*	**95 *in* 100**	**5 *in* 100**	☐
21)	*Outcome of Game*	☐	*Or*	**100 *in* 100**	**0 *in* 100**	☐

Page 1