Will You Commit to Trust? Investigating the role of commitment statements in promoting trustworthy behavior

Matthew Gildea, Cameo Hazlewood, Chloe Lin, Julia Moore, Tal Tahori, Vaidehi Uberoi, Fiona Yang

University of Pennsylvania

ABSTRACT

Previous research has extensively studied the dynamics of trustworthiness between two players in a trust game. While some studies have demonstrated the effectiveness of commitments in establishing trust, the extent to which commitment devices are an effective mechanism for driving desired behavior needed additional exploration. This experiment conducted at the University of Pennsylvania builds on the existing trust game literature by exploring the role of commitment statements in instilling trustworthy behavior. Specifically, we investigate whether agreeing to a commitment statement to act in a trustworthy manner before playing repeated online trust games will enhance the trustworthy behavior of the second mover. We used an online trust game interface paired with instructional envelopes that a) gave the second mover the option to commit to trustworthy behavior or b) required the second mover to commit to trustworthy behavior and subsequently gave the option to recommit to trustworthy behavior. The results did not show a significant improvement in trustworthy behavior during the game. However, the study found that participants who were required to agree to a mandatory initial commitment statement were more likely to self-select to recommit to the statement at the midway point. This provides evidence for the status quo bias, a behavioral concept that explains our irrational preference for the current state of affairs.

Keywords: trust, trustworthiness, commitment, contracts, compliance

INTRODUCTION

Trust is the foundation for most social and economic interactions and a basis around which human relationships revolve. Studies have found a correlation between willingness to behave in a trustworthy manner and improved economic outcomes (Arrow, 1972; Fukuymam, 1995). Lawyers and business partners use trust and commitment contracts to establish trustworthiness to enhance engagement during business deals. Policymakers may aim to change collective behavior by improving commitment and trust among individuals. For example, a campaign to decrease the amount of meat consumed in the United States might try to use personal commitment contracts to help individuals commit to change this behavior. Similarly, when it comes to charitable giving, nonprofits may need to build trust in society to establish reciprocal trustworthy relations. Trust is an important facet in building relationships and in expecting a favorable outcome in several domains. It is therefore important to understand the behavioral mechanisms that can be applied to build and enhance the level of trustworthiness between dyadic relationships.

Previous research has extensively studied the dynamics of trust and trustworthiness between two players in an economic trust game. Research has been done to study the effectiveness of variables such as observability, contracts with punishment, and pre-binding commitments on decision making and the establishment of trust between players (e.g. Bracht & Feltovich, 2008). Malhotra and Murnighan (2002) found that binding contracts make the interacting parties attribute others' cooperation solely to the constraints imposed by the contract rather than to the individuals themselves. Their research suggests that non-binding contracts led to personal attributions for cooperation and thus, may provide an optimal basis for building interpersonal trust in a variety of situations. Schweitzer, Ho, and Zhang (2018) suggest that second movers are compliant when they know in advance that their actions will be monitored and recorded but exploited their counterparts when they know in advance that they would not be recorded. This research leads to implications for the follow-through of the desired behavior change. Additionally, Rogers, Milkman, and Volpp (2014) researched pre-commitment devices as a tool to change behavior, which showed that the use of commitment devices may be an effective mechanism. However, more research is needed to determine the direct effects on behavior.

The study reported here seeks to build on the existing literature by exploring the role of commitment statements in instilling trustworthy behavior between two individuals. We use the standard two-player economic trust game developed and popularized by Berg, Dickhaut, and McCabe (1995) and focus specifically on the behavior of the second mover in a series of six repeated games. We aim to study the effect of incorporating mandatory and/or voluntary

commitment statements at different time points during the six rounds. Depending on the treatment condition, second movers were presented with no commitment statements (Control), a midway point *voluntary* commitment decision (Treatment 1), or an initial *mandatory* commitment statement and a midway point *voluntary* commitment decision (Treatment 2). By studying the decisions and behaviors of the second movers, we aim to determine whether commitment statements play a significant role in trustworthiness. To analyze this, we will measure the amount returned each round by the second mover recorded as a percentage of the amount received from the first mover and whether each round's return is trustworthy (which we define as returning equal or more points than initially sent). We predict that second movers who agree to a commitment statement will exhibit greater trustworthiness during the trust games.

We contribute to the gap in the literature by exploring how agreeing to commit to trustworthy behavior impacts the decisions of second movers in repeated trust games. We found that neither requiring participants to agree to an initial commitment statement nor their voluntary decision to choose to agree to a midway point commitment statement has a significant impact on their trustworthy behavior in subsequent rounds of trust games. However, second movers who agree to a mandatory initial commitment statement were more likely to self-select to recommit to the statement at the midway point. These findings support research showing that individuals are prone to the status quo bias, or the behavioral preference for the current state of affairs or a previously made decision (Samuelson & Zeckhauser, 1988).

In this paper, we provide a review of the literature on trustworthy behavior and commitments before presenting our research questions, methods and design, hypotheses, and results of the experiment. We conclude with a discussion of our findings, the limitations of our research, challenges faced, and the broader policy, business, and behavioral implications of the reported results.

LITERATURE REVIEW

Over the last couple of decades, researchers have explored the role of measuring trust as an economic model of behavior (meta-analysis by Johnson & Mislin, 2011). First introduced by Berg, Dickhaut, and McCabe (1995), the now popular economic trust game has become a valid model for measuring *trust* and *trustworthiness*. This simple two-player game involves a sequential exchange in which the first mover decides how many of their endowed points or dollars they will send to an

anonymously paired partner knowing that the amount they send will be tripled by the experimenter. The second mover then decides how many of the now tripled points or dollars they received to send back to the first mover. The amount sent by the first mover is understood to capture *trust* or "a willingness to bet that another person will reciprocate a risky move (at a cost to themselves)'' and the amount returned by the second mover is understood to capture *trustworthiness* (Camerer & Fehr, 2003, p. 85).

Research by Levine, Bitterly, Cohen, and Schweitzer (2018) used the trust game to link specific personality traits and situational dispositions to interpersonal trustworthiness. They find that guilt-prone individuals are more trustworthy than individuals who are low in guilt-proneness, but they are not universally more generous. This may relate to potential self-selection into agreeing to commit or not commit. The personality type of individuals may influence them to commit to being trustworthy and further reflect in their behavior.

Wilkinson-Ryan (2012) found that signing a contract during the trust game in which punishment is associated with defective behavior is a strong indicator of how trustworthy an individual behaves during the game. Although punishment can be a strong signal of behavioral compliance, we decided to measure the effectiveness of commitment statements on behavior without the mediating factor of punishment. We were encouraged to study the impact of commitment statements as a positive reinforcer in itself and not bundle it with the negative reinforcing elements of punishment. Additionally, Ederer and Schneider (2019) show that the passage of time does not negatively affect trust, trustworthiness, and cooperation, even after three weeks of time had passed. This is support for our study that was conducted in the span of about ten minutes; trustworthy behavior, if any, would have remained, which shows potential external validity.

Malhotra and Murnighan (2002) and Simpson and Eriksson (2009) used the trust game to measure the effect of contracts on trustworthy behavior. Malhotra and Murnighan found that binding contracts, or contracts containing punishment conditions, make the interacting parties attribute another person's cooperation to the constraints imposed by the contracts alone, rather than to the individual. Based on this finding, we chose not to inform first movers of our study that their partners would be agreeing to a commitment statement. We did not want the trustworthy effect to fail and influence the sender to believe that the amounts being sent are a result of a "contract" and not a natural playing behavior, as it had in the aforementioned studies. Furthermore, research by Schweitzer et al. (2018) suggests that receivers are compliant when they know in advance that their actions will be recorded but are more likely to exploit their counterparts when they know in advance that their actions would

not be recorded. For this reason, as the experimenters, we made sure that we were not influencing participants' perceptions of being observed by conducting the study in separate rooms than where the experimenter was stationed.

Research on pre-commitment devices as a tool to change behavior shows that the use of commitment devices may be an effective mechanism of behavior change (Rogers, Milkman, & Volpp, 2014). A study by Baca-Motes, Brown, Gneezy, Keenan, and Nelson (2012) studied engagement in environmental behavior change as measured by hanging one's towel in a hotel as a function of a commitment device in the form of a pin. They found that people who decided not to commit to behaving in an environmentally friendly behavior (no pin) were less environmentally friendly than the control group (no manipulation), which is evidence for potential self-selection of commitment.

RESEARCH QUESTIONS

Our overall research objective was to investigate the impact of initial mandatory commitments on trustworthy behavior in repeated trust games and the impact of self-selection to agree or not agree to midway point commitments on trustworthy behavior in subsequent repeated trust games. We designed our experiment to test the following research questions specifically:

> **Research Q1a:** Will participants who agree to a mandatory initial commitment statement show greater trustworthy behavior than those who are not required to agree to an initial commitment statement?

In research question 1a, the independent variable is an agreement to a mandatory initial commitment statement; the dependent variable is the second mover's trustworthy behavior in Trust Games rounds 1-3.

> **Research Q1b:** Will participants who agree to a mandatory initial commitment statement be more likely to self-select to agree to the new midway point commitment statement than those who are not required to agree to an initial commitment statement?

Research question 1b also includes an agreement to a mandatory initial commitment statement as the independent variable. It incorporates participants' likelihood to agree to a midway commitment statement as the dependent variable.

> **Research Q2:** Will participants who self-select to agree or re-agree to a midway point commitment statement show greater trustworthy behavior in subsequent trust games than those who did not agree or re-agree?

Research question 2 examines the independent variable, self-selection to agree or re-agree to a midway commitment statement. This would be a predictor of the dependent variable, the second mover's trustworthy behavior, in rounds 4-6.

> **Research Q3:** Will participants who are introduced to a midway point commitment statement and choose to commit show the greatest trustworthy behavior across all treatments? Will participants who are introduced to a midway point commitment statement and choose NOT to commit show the lowest level of trustworthy behavior across all treatments?

Lastly, research question 3 examines the relationship between the independent variable, exposure (or non-exposure) to midway commitment statement, and the dependent variable, the second mover's trustworthy behavior in rounds 4-6.

EXPERIMENTAL DESIGN

Methodology

We conducted a laboratory experiment to measure the behaviors of individuals playing six repeated rounds of the trust game. Participants were randomly and anonymously paired with another participant. In each dyad, one participant was assigned the role of first mover (Player A) and the other the role of second mover (Player B). The players kept their role throughout the entire experiment. Each of the six rounds was a standard two-player trust game (Berg et al., 1995) where the amount sent by the first mover is tripled. The second mover then decides how many of the tripled points to send back to the first mover (see ***Figure 1***).

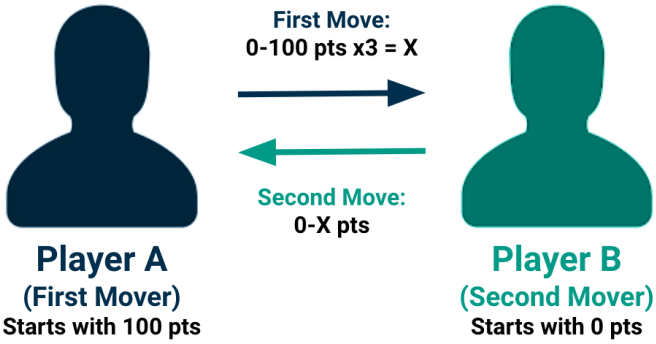


Figure 1. Standard 2-player trust game. To start, Player A receives 100 points, Player B receives nothing. Player A can send some or all of his 100 points to Player B. Before Player B receives those points they are tripled. Once Player B receives the tripled points, he can decide to send some or all of his points back to Player A.

We used a 2 x 3 design to test different scenarios of commitment mechanisms designed to improve Player B's trustworthy behavior in the trust games. We measured the impact of commitment statements on Player B's behavior at two different time points. First, immediately before any trust games had been played and second, at a midway point in the trust games. For the initial commitment statement, the experimenters mandated agreement by Player B because we wanted to measure if participants who agree to a mandatory initial commitment statement show greater trustworthy behavior than those who are not required to agree to an initial commitment statement. The midway point commitment statements were always presented as a self-select decision in our treatment groups to which Player B could choose to agree or not agree. This design allowed us to analyze the following questions. First, would participants who agree to a mandatory initial commitment statement be more likely to self-select to agree to a new midway point commitment statement than those not required to agree to an initial commitment statement. Second, would participants who self-select to agree or re-agree to a midway point commitment statement show greater trustworthy behavior in subsequent trust games than those who did not agree or re-agree. Third, would participants who are introduced to a midway point commitment statement, and choose to commit, show the greatest trustworthy behavior across all treatments.

Each dyad was assigned to either one of two treatment conditions or the control condition. In the Control (C) condition, pairs play six consecutive rounds without exposure to an initial commitment statement or a midway point commitment statement. In Treatment 1 (T1), Player B was exposed to a midway point commitment decision after Rounds 1-3 in which they had to self-select to agree or not agree to a commitment statement stating, "I agree to act trustworthy in the

following interaction" in Rounds 4-6. In Treatment 2 (T2), Player B was required to agree to a mandatory initial commitment statement before Rounds 1-3 stating, "I agree to act trustworthy in the following interactions." They were then exposed to a midway point recommitment decision after Rounds 1-3 in which they had to self-select to agree or not agree to recommit to trustworthy behavior in Rounds 4-6. ***Table 1*** summarizes our 2 x 3 (no mandated initial commitment vs. mandated initial commitment x no midway commitment decision vs. midway self-selected commitment vs. midway no self-selected commitment) design.

	First Movers		**Second Movers**	
Trust Games	**Before Rounds 1-3**	**Before Rounds 4-6**	**Before Rounds 1-3**	**Before Rounds 4-6**
C	No commitments		No commitments	
T1	No commitments		No mandatory commitment	Self-select to commit
T2	No commitments		Mandatory commitment	Self-select to recommit

Table 1. Summary of Control and Treatment Arms.

Procedure

We conducted a laboratory experiment on the University of Pennsylvania campus. Data collection ran between November 12, 2019, to December 2, 2019. 220 undergraduate and graduate students were recruited from the University of Pennsylvania to participate in the study. For each dyad, we recruited participants one at a time and separated them into different rooms to ensure their identities were anonymous to one another. In the room, each participant was provided with a computer, three closed envelopes, and a pen. The recruiter instructed the participant to open Envelope 1, which would provide initial instructions regarding the experiment (see Appendix C for *Envelope 1 Instructions*). Participants were told they would be given one raffle ticket to win a $100 Amazon gift card for their participation and an opportunity to win additional raffle tickets based on their decisions and their partner's decisions during the following economic interactions. For T2 second movers only, Envelope 1 instructions were modified to include a mandatory

commitment statement on the first page. These participants were instructed to rewrite a commitment stating "I agree to act trustworthy in the following interactions" and sign their initials to commit to trustworthy behavior toward their partner in the experiment. We defined trustworthy behavior in the instructions as "actions one party takes that fulfill both their own interests and the interests of the other party." T2 first movers were not made aware of the commitment statement their partner was forced to agree to. The instructions in Envelope 1 then walked all participants through the login procedures with their unique game codes on economic-games.com — our online source for trust games powered by *oTree*. The trust game interface provided onscreen instructions for playing the game, followed by an understanding question as a comprehension check. Participants played three rounds of trust games with their partner and then were instructed to open Envelope 2.

Envelope 2 told participants they had completed the first half of the study and gave instructions to login to the second half of the study again using a unique game code to access online trust games through economic-games.com (see Appendix C for *Envelope 2 Instructions*). Participants were informed they would maintain the same role (Player A or Player B) and would be playing with the same anonymous partner. For T1 second movers, Envelope 2 instructions were modified to include a voluntary commitment statement on the first page. These participants were asked to decide whether they wanted to commit to trustworthy behavior toward their partner in the remainder of the experiment. We defined trustworthy behavior in the instructions as "actions one party takes that fulfill both their own interests and the interests of the other party." To commit, they were instructed to rewrite a commitment stating "I agree to act trustworthy in the following interactions" and sign their initials. Those who chose not to commit were simply instructed to leave the space blank. For T2 second movers, Envelope 2 instructions were also modified to include a voluntary re-commitment statement on the first page. We redefined trustworthy behavior and asked those who chose to recommit to rewrite the same commitment statement and sign their initials while those who chose not to recommit were told to leave the space blank. Neither T1 or T2 first movers were informed of their partner's exposure to a commitment decision or the outcome of that decision. All participants again followed the trust game interface onscreen instructions and played three additional rounds of trust games with their partner. At the end of the three trust games, participants were instructed to open Envelope 3.

Envelope 3 contained a post-experiment questionnaire that asked all participants their age, gender, and current level of schooling (see Appendix C for *Envelope 3 Instructions*). We also asked

participants if they have had past experience with economic trust games like the ones in this study. Finally, participants were asked to provide their email so we could contact them if they won the $100 Amazon gift card. After completing the questionnaire, participants were instructed to leave all materials inside the room and meet the experimenter outside who would inform them of the total number of raffle tickets they earned during the online experiment.

Figure 2 below provides an overall scheme of the experimental procedure. The envelope instructions provided to participants and screenshots of the online game interface can be found in *Appendix B* and *Appendix C*.

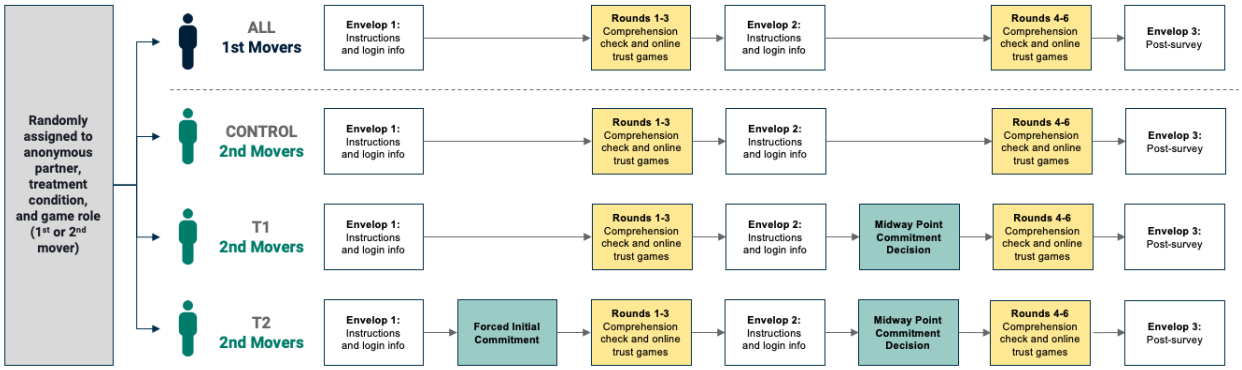


Figure 2. Experimental Procedure Schematic.

Pilot Study

To reflect all the procedures of the main study and validate the feasibility of design methods, we ran an initial pilot study that included 16 dyads (32 individual participants). Half of the dyads were randomly assigned to T1 (N=8), and half were randomly assigned to T2 (N=8). Of these, we were able to include 11 dyads in our final analysis (N=6 T1 and N=5 T2). By testing different recruiting locations around the campus, we found libraries and public study areas are best for the purpose of making sure the inclusion and exclusion criteria of the target population are met (i.e., they are current students at University at Pennsylvania as the inclusion criteria). We made sure to run the pilot study with all members of the research team to ensure proper training and execution consistency throughout all data collection sessions.

The pilot study also allowed us to test the measurement instrument and identify minor adjustments that needed to be made to the participant instructions to ensure proper understanding and adherence to the experimental procedures. We learned of additional details of our experimental procedures that proved to be unclear or not obvious enough and resulted in participants' confusion with or in-adherence to the instructions. Therefore, we made the following minor, yet advantageous revisions to the participant instructions for the main study. First, we added a screenshot of the

online interface to clarify the login process. Second, we found that some participants overlooked content printed on double-sided materials, and we fixed this issue by adding clear instructions to "please see reverse side" on the bottom of relevant pages to ensure participants saw all of the instructions. Third, we found that while the instructions stated that T2 second movers "must agree" to the initial commitment statement, some failed to do so. We addressed this problem by bolding, underlining, and italicizing the phrase "you must agree" to add additional emphasis on the mandatory action. Finally, we added a box around the actual commitment statements in both T1 and T2 instructions to make them more prominent, adding emphasis on the task or decision. These slight adjustments to the participant instructions were effective in terms of creating more clarity and driving adherence in our main data collection.

Other than the issues mentioned above, the preliminary data aligned with the pattern we hypothesized, and we used it to conduct a preliminary power calculation to get the expected number of participants we would need to recruit to achieve the desired statistical power for our design.

HYPOTHESES

According to Kellner, Reinstein & Riener (2019), pre-commitment improves prosocial behavior (charitable giving). Since behaving in a trustworthy way is a form of prosocial behavior, we could hypothesize that agreeing to a pre-commitment would lead to more trustworthy actions on participants' end, which is part of Treatment 2. The reason why pre-commitment promotes prosocial behavior is further explored in Breman's (2011) study on the psychological "warm glow" that occurs when the committing takes place. Such a positive psychological effect can account for the increasingly prosocial and therefore more trustworthy behaviors by participants who are instructed to pre-commit. This led to our first hypothesis:

> **H1a:** Participants who agree to a mandatory initial commitment statement (T2) will show greater trustworthy behavior in trust games than those who are not required to agree to an initial commitment statement (C/T1).

Samuelson & Zeckhauser (1988) studied the effect of status quo bias in decision making when economics games are used. They found that individuals are most likely to maintain their previous decision, or disproportionately maintain the status quo, in decision-making tasks. This assertion led to our next hypothesis:

> **H1b:** Participants who agree to a mandatory initial commitment statement (T2) will be

more likely to self-select to agree to the new midway point commitment statement than those who are not required to agree to an initial commitment statement (C/T1).

According to Lewis, Goetz, Schoenfield, Gordon & Griffin (1984) and Berscheid, Boye & Darley (1968), individuals who commit voluntarily to contracts are usually motivated to do so because of either a tangible reward or a consequence. Additionally, Andreoni and Serra-Garcia (2019) show that providing participants with the option to self-select into pledging a gift increases their charitable giving. Hence, we assumed that participants who would voluntarily self-select into committing without any external reward or consequence would do so because of an intrinsic motivation to commit and hence would behave in a more trustworthy behavior than those who did not agree to commit to the statement voluntarily. This led to the development of our final two hypotheses:

H2: Participants who self-select to agree (T1B) or re-agree (T2B) to a midway point commitment statement will show greater trustworthy behavior in subsequent trust games than those who did not agree (T1A) or re-agree (T2A).

H3: Participants who are introduced to a midway point commitment statement and choose to commit (T1B and T2B) will show the greatest trustworthy behavior across all treatments. Participants who are introduced to a midway point commitment statement and choose NOT to commit (T1A and T2A) will show the lowest level of trustworthy behavior across all treatments.

ANALYSIS AND RESULTS

Descriptive Statistics

According to our research question and hypotheses, the main subject of interest is the second mover in each game. In ***Table 2***, an overview of our study with the average of the incidence of trustworthy behavior, number of observations in each subgroup, demographic information such as the gender distribution, average age, and the number of participants with previous trust game experience is presented. Explanation in detail about the inference of all the summary statistics are included in the following sections.

Group	Phase	Self-Selection	Incidence of Trustworthy Behavior Group Mean	Size	Gender	Age	# of Experience of Trust Game
Control	1	-	0.5333333	30	30 Males	22.6	16
	2	-	0.5111111	30	30 Males	22.6	16
Treatment 1	1	-	0.3645833	32	15 Males, 7 Females	22.40741	13
	2	A	0.4444444	18	8 Males, 10 Females	22.86667	6
		B	0.5952381	14	7 Males, 7 Females	21.83333	7
Treatment 2	1	-	0.4666667	30	16 Males, 4 Females	22.77778	10
	2	A	0.4285714	7	4 Males, 3 Females	23.66667	2
		B	0.6521739	23	12 Males, 11 Females	22.52381	8

Table 2. Second Mover Summary Statistics.

Statistical Tools & Methods

We used Google Sheets to record our primary data, so the statistical tools involved to help clean and analyze the data included Excel and R Studio. We converted the excel sheet into a CSV file so that it could be easily read in R studio.

Using the results from valid 12 dyads of participants collected from the pilot study, we ran a power analysis on G*Power to determine the sample size we needed for each treatment group to achieve 0.80 power with a p-value of .05. Data showed the Treatment 1 group mean incidence of positive returning behavior for second-mover was 0.7777778 with a standard deviation of 0.2721655; the Treatment 2 group mean incidence of positive returning behavior for second-mover

was 0.83333333 with a standard deviation of 0.2788867. Therefore, we ideally needed to obtain 406 independent pairs per treatment group, which was 1218 pairs in total.

Due to restrictions on time and resources, we considered the possibility that we would not reach the sample size suggested by that calculation. Therefore, we aimed to collect at least 30 pairs for each treatment, leading to 90 pairs in total, which would allow us to achieve a power of 11.65%[1].

Table 3 includes the description of variables that we collected during the experiment in order and those we created during data cleaning for the ease of analysis.

Data	Type	Description
Treatment Group	Categorical	Control, Treatment 1 and Treatment 2
Role	Binomial	1 stands for first mover, and 0 stands for second mover
Points sent and received	Ordinal	Number of points sent and received during each round of the games. There are 2 games and 6 rounds in total
Percentage returned	Ordinal	Percentage of points returned during each round of the games compared to the points received
Cumulative Points	Ordinal	Number of cumulative points in hand by the end of Game 1 and Game 2 respectively, and the combination of the two
Average percentage returned	Ordinal	Average percentage of points returned during Game 1 and Game 2 respectively, and the combination of the two
Midway point commitment	Binomial	1 stands for agreeing to commitment statement, and 0 stands for not
Raffle tickets earned	Ordinal	Number of raffle tickets earned according to pre-decided conversion rate (1 for showing up)
Demographics	Categorical	Age, gender, school level, trust game experience, email
Trustworthyincid_phase1	Ordinal	Incidence of trustworthy returning behavior (defined as returning equal or more than received) in Game 1

[1] For a post hoc analysis, see *Appendix A*

Trustworthyincid_phase2	Ordinal	Incidence of trustworthy returning behavior (defined as returning equal or more than received) in Game 1

Table 3. Descriptions of variables.

The data of interest are the binomial outcome of midway point commitment. The incidence of trustworthy behavior measured by calculating the ratio of times out of 3 rounds in each game that the second movers returned as much as they received.

As a result, a total of 110 dyads played our modified trust game. Randomization for recruiting of each treatment group and role of first or second mover was realized by online random number generator website, and our criteria for outlier was defined that first mover sent 0 points in every round of either of the two phases (rounds 1-3 and rounds 4-6). Vetting out invalid data and outliers, we reached a number of 92 dyads eligible for data analysis with 30 dyads in Control group, 32 for Treatment 1 group, and 30 for Treatment 2 group.

The Nonparametric Wilcoxon test was the primary statistical test we used in the study since no assumption about the underlying distribution would be needed. Linear regression models were also useful to present the relationship between several demographic variables and determine the important ones that could contribute to behavioral change.

Results

Result 1a*: Participants who agreed to a mandatory initial commitment statement did not show greater trustworthy behavior in trust games than those who were not required to agree to an initial commitment statement.*

We found no significant evidence supporting hypothesis 1a. To test this hypothesis, we conducted a nonparametric Wilcoxon test among the three treatment groups. We analyzed the second mover's behavior in regards to the amount returned to the first mover and compared the proportion of amounts returned that are considered trustworthy across the first three rounds of each treatment group. Thus, we found that second movers who were required to agree to a commitment statement in Treatment 2 did not show significantly greater trustworthy behavior than second movers who were not introduced to any commitment statement in Treatment 1 and the Control.

The resulting p-values for the test between Treatment 2 and Treatment 1 and the test between Treatment 2 and Control were 0.239 and 0.1004, respectively.

	T1	T2	Control
Dyads of participants	n=32	n=30	n=30
Incidence Group Mean	0.365 (0.409)	0.467 (0.388)	0.533 (0.425)

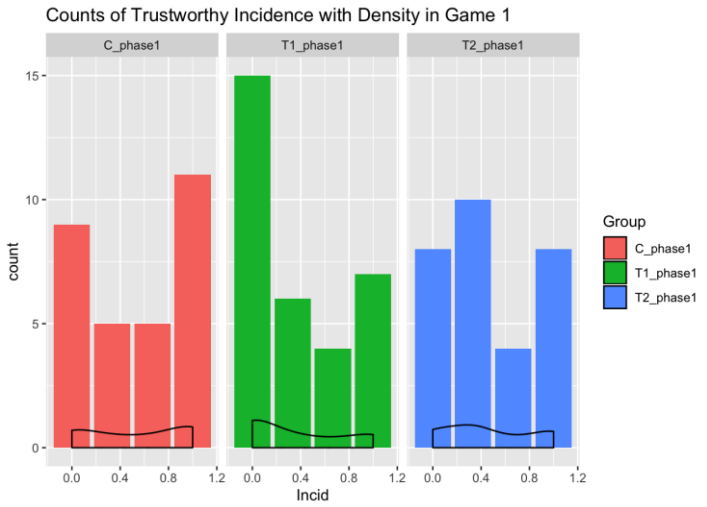


Figure 3. Barplots of the counts of trustworthy incidence with density distribution in Game 1. Incidence of 0 (first bar in each segment) means the second movers did not return once more than they received during all three rounds in Game 1. Incidence of 1 (last bar in each segment) means the second movers returned more than they received every time during all three rounds in Game 1.

As shown in ***Figure 3***, the barplots of the counts of trustworthy incidence with density distribution over each group in the first game did not appear very different, visualizing the similar trustworthy behavior across the groups. To further explore if there were other factors that could differentiate this trustworthy behavior, we separated each treatment group by gender and by previous experience of trust game and conducted the tests again. However, the results remained statistically insignificant, indicating regardless of gender and awareness of the experiment, participants did not behave in a more trustworthy manner even though they agreed to in a commitment statement.

Result 1b*: Participants who agreed to a mandatory initial commitment statement would be more likely to self-select to agree to the new midway point commitment statement than those who were not required to agree to an initial commitment statement.*

We found significant evidence supporting hypothesis 1b. To test this hypothesis, we conducted a 2-sample t-test using a Bernoulli distribution to indicate whether participants agreed to the midway commitment statement or not. As shown in ***Figure 4***, 14 out of 32 participants in the second mover group in Treatment 1 voluntarily agreed to the first-time introduced commitment statement, and the rest of the 18 participants in the second mover group did not agree. However, in Treatment 2, in which it was mandatory for participants to commit, 23 out of the 30 participants in the second mover group recommitted, and only 7 of them opted out. We thus provide evidence showing that participants who agree to a mandatory initial commitment statement are more likely to self-select to agree to the new midway point commitment statement than those who are not required to agree to an initial commitment statement. A p-value of 0.007 confirmed the statistical significance in the midway commitment statement choice between the two treatment groups.

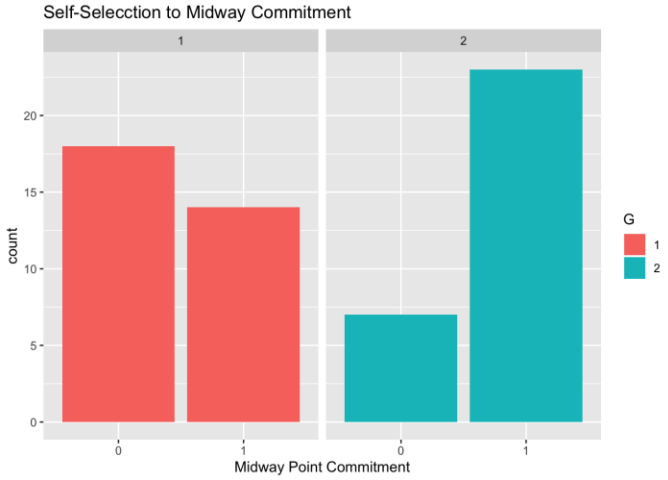


Figure 4. Barplot of self-selection to midway commitment statement. 1 stands for signing initials, while 0 stands for not with the left segment displaying Treatment 1 second movers, and the right segment displaying Treatment 2's.

Similar to the above procedure, we also tested if the willingness to recommit to the midway point commitment statement significantly varies based on gender difference and experience with the trust game. We again found significant results in a comparison between male or female participants, indicating gender was not the deciding variable in midway point selection. However, statistical significance did not appear in a test only taking participants with trust game experience

into account (p-value = 0.147), suggesting participants who were familiar with the game were less likely to recommit in Treatment 2 as we expected. This result was not too surprising because participants with some knowledge about the game might have a winning strategy in mind and, therefore, they did not want to be restricted by certain behavioral patterns. However, in general, our main significant result suggests that guiding people to agree on some commitment statements might increase the chance to have the same people repeat the action in the future.

Result 2*: Participants who self-selected to agree or re-agree to a midway point commitment statement would not show greater trustworthy behavior in subsequent trust games than those who did not agree or re-agree.*

To test our hypothesis 2 in terms of trustworthy behavior in Game 2 after different midway point self-selection, we conducted Wilcoxon tests within each of Treatment 1 and Treatment 2 to see if the incidence of trustworthy behavior was more likely to take place among those who agreed to the commitment. However, the p-value for the test between participants committed to the statement (T1B) and those who did not (T1A) in Treatment 1 was 0.3008, and for the test between participants recommitted to the statement (T2B) and those who opted out (T2A) in Treatment 2 was 0.1915. Both tests failed to prove agreeing to a commitment statement at the second stage could build participants' trustworthiness. These results seemed counterintuitive because we would expect to see commitment statement intervention have some influence over trustworthy behavior. However, a likely explanation could be that incentives were not sufficient to drive to follow the agreement they made without certain punishment for deviation in the game.

	T1A	T1B	T2A	T2B
Dyads of participants	n=18	n=14	n=7	n=23
Incidence Group Mean	0.444(0.443)	0.595 (0.417)	0.429 (0.418)	0.652(0.355)

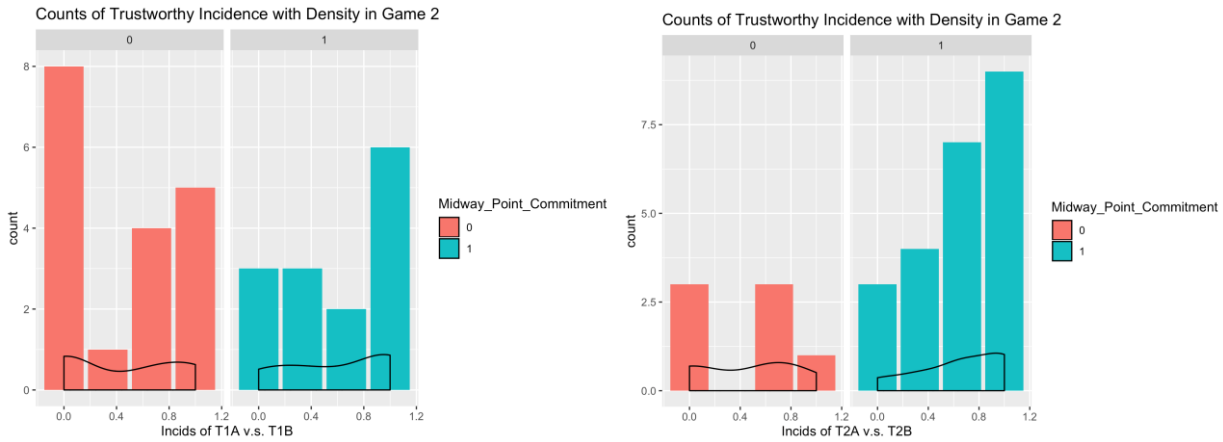


Figure 5. Barplots of the counts of trustworthy incidence with density distribution in Game 2. Orange bars represent second movers who did not agree to the midway commitment statement, and blue bars represent those who agreed. The left two barplots stand for Treatment 1, and the right two stand for Treatment 2.

Although the commitment statement was not significant in promoting trustworthy behavior according to statistical tests, it yielded some intriguing observations. It is particularly interesting to look at the counts and density distribution for each group, as shown in ***Figure 5***. The counts of each incidence level might not be meaningful since subgroups after self-selection were apparently different in size, while the density distribution over participants who agreed and signed their initials (colored blue) was more skewed to the right (higher incidence to return trustworthily) than that of participants who did not agree or sign their initials (colored orange) in both Treatment 1 and Treatment 2. This tendency, to some extent, exhibited more occurrence of trustworthy returning behavior in our study. In addition, the group means of 0.652 and 0.595 were also higher than 0.429 and 0.444, respectively, confirming the direction of influence in our hypothesis. Perhaps due to the small sample size, especially with one subgroup consisting of only 7 observations, the numbers from the tests failed to reflect the true findings explicitly.

In addition to within-group tests, we conducted an analysis to compare the results of participants who re-committed to the commitment (T2B, colored blue) with those who were first brought to the intervention (T1B, colored blue). This between-group test can help us understand if subsequent commitment was significantly more powerful in inducing trustworthiness compared to a single commitment. Again, the p-value of 0.7675 suggests no statistical significance, which could potentially undermine the effectiveness of the commitment statement in regards to trustworthiness.

Additional statistics of tests separating groups by gender and experience with trust game did not show any significance. So in all, hypothesis 2 was not supported by our experiment results.

Result 3*: Trustworthy behavior was similar across all treatment groups during Game 2.*

In hypothesis 3, we expected to see that the level of trustworthiness, as determined by the amount returned by the second mover, is highest among participants who were introduced to a midway point commitment statement and choose to commit. Consequently, we expected trustworthiness to be the lowest among participants who were introduced to a midway point commitment statement and choose NOT to commit. We therefore anticipated trustworthiness to be as followed: T2B > T1B > C > T2A > T1A). To test this, we could either use the Bonferroni test or conduct a pairwise comparison for each of the groups. We used the latter method to comply with the pre-registration plan and conducted 10 Wilcoxon tests in total (4 of which were previously conducted to test hypothesis 2).

Corresponding to outcomes seen in the third result, none of the comparisons across these five treatment groups yielded a significant p-value for Game 2. Therefore, we were unable to rank them against each other.

Our data analysis suggests that commitment statements do not significantly increase trustworthy behavior. However, there are several possible explanations that could potentially explain the null results. Notably, we suggest that the design of the experiment, the range of data confined to the second mover, and others, could serve as possible reasons. Additional findings and noteworthy observations will be discussed in the following *Discussion* and *Appendix A*.

DISCUSSION

In the scope of this paper, we found three overarching results. The purpose of this study was to evaluate the effect of commitment statements on influencing trustworthy behavior. Our primary hypotheses (H1a & H3) assured that participants who agree to a commitment statement would act more trustworthy than those who do not. However, we found no significant statistical results to support these claims and, thus, failed to reject the null hypothesis that commitment statements do not increase trustworthy behavior. This is true under both hypothesized situations. First, there is no significant difference in trustworthy frequency between participants who agreed to the mandatory commitment statement at the beginning of the study (T2) and those who did not (Control & T1). Secondly, participants who self-selected to agree to the midway commitment statement (T1B & T2B) showed no significant increase in trustworthy behavior than those who did not agree (Control, T1A, and T2A). Overall, these null results imply that a commitment statement

alone does not have a significant impact on an individual's decision to act in a trustworthy manner.

While this is a null finding, its importance should not be understated. Specifically, this lack of significant differences between participants who agree to commitment statements and those who do not invalidates one of the bases underlying the motivation of this study. That is that observers assumed that individuals only acted trustworthy because they were under contract (Malhotra & Murnighan, 2002). Our results suggest that since participants do not act more trustworthy under commitment statements, this assumption is irrational and possibly detrimental to the creation of interpersonal trust. Additionally, our results oppose those found by Baca-Motes et al. (2012), which illustrated that a brief, specific commitment at a hotel check-in increased environmental friendliness (using towels for multiple days) by 25%. As our study found no significant difference between those under commitment and those not, our results challenge the efficacy of these effects. However, it should be noted that these differences could be influenced by a variety of factors other than the true effect of commitment statements. Firstly, the original study offered those who committed a lapel pin for their commitment, which could factor into their behavior through the reciprocity bias in which individuals feel compelled to return the favor (in this case the gift of the pin) by adjusting their behavior to meet expectations. Secondly, Baca-Motes et al. performed a field experiment wherein participants' behavior was secretly recorded under a situation with little incentive to break their initial commitment (reusing towels), while out study was built around a trust game that incentivizes participants to act selfishly to earn a larger amount of tickets. It could be that this inherently more point incentivized game crowds out the effect by pushing all participants, including the commitment groups, to act less trustworthy and thus, less in line with their commitment.

When examining participants' likelihood to self-select to agree to a midpoint commitment statement, we found significant evidence supporting hypothesis H1b. Our analysis shows that our findings are consistent with the literature as participants who agree to a mandatory initial commitment statement were more likely to agree to a new midway point commitment statement than those who were not required to agree to one initially. Here, evidence from behavioral sciences is particularly relevant as these findings are consistent with the status quo bias. According to the status quo bias, people prefer to remain consistent and stand by their original decisions (Samuelson & Zeckhauser, 1988).

Our analysis for hypothesis H2 shows no significant evidence that self-selection yields greater trustworthy behavior. Interestingly, these results go against our expectations and refute the

general trend that the current literature shows. Unlike the presumption that self-selection encourages greater giving (Andreoni & Serra-Garcia, 2019), we found that participants who voluntarily agreed to re-agreed to a commitment statement have not shown significantly higher trustworthiness than participants who did not. Building on literature that draws a connection between voluntary behavior and tangible rewards (Lewis et al., 1984; Berscheid, 1968), these results can be explained due to the low salience of consequences and tangible rewards in regards to the commitment.

Our research provides relevant theoretical and practical contributions to the fields of behavioral science, policy, and business. We provide significant evidence to the field of behavioral science by providing substantial support to the status quo bias and suggest that initial commitment statements may serve to preserve behavior change. These findings are particularly relevant for the development of behavioral policies and suggest that policymakers may change collective behavior by introducing initial and midpoint commit statements to behaviorally inspired interventions. Additionally, these findings are also applicable to various business interactions. We argue that incorporating initial and midpoint commitment statements in contracts and business agreements may enhance future engagement in business dynamics.

Overall, this study demonstrates that commitment statements without punishment have no significant effect on an individual's likelihood of acting trustworthy; however, this study does find that those who commit are more likely to recommit to another agreement. This refutes previous literature that suggests people do increase giving under contracts or at least assume that others under contract display this effect. Moreover, these findings support research that cheap talk is not effective at changing behavior as well as reinforce the concept of status quo bias under the context of commitment statements.

Limitations and Future Research

It should be noted that our study contains multiple limitations. Firstly, our goal was to gather enough observations per treatment to achieve 80% power (95 dyads per treatment based on power calculations from previous papers published on trust games), yet due to limited resources and time, we were only able to gather roughly 33% of our desired sample size. This does not nullify our findings but means that all results should be taken with a grain of salt as there is above a 50% chance that these findings are not robust and may have been acquired by chance alone. Secondly, nearly 30% of participants that were instructed to agree to the initial mandatory commitment

statement (T2) failed to comply. This could either be due to a lack of focus and comprehension of the instructions or an unwillingness to agree to or sign a contract at all. Either way, these factors could be correlated to a third (or more) unknown variable that affects trustworthy behavior. Thus, while we ran more dyads in Treatment 2 to reach an N=30, these participants may be biased by their understanding and willingness to commit and not be representative of the true population we sought to test. Additionally, since our experiment consisted of two separate, 3-round trust games, the end-game effect may largely crowd out natural fluctuations in trustworthy behavior. ***Figure 6*** illustrates that both rounds three and six had significant drops in trustworthy behavior frequency which means that this effect was present in a third of our data. While this effect may limit the number of useful data points we collected the effect appears across groups and therefore likely evens out amongst treatments limiting the negative trustworthy bias this effect produces.

We further believe that our payment scheme did not provide enough incentive to act dishonestly. As we only had $100 in the form of an Amazon gift card to give out, the expected value of each individual's move by each participant was negligible (roughly 7 cents re round). This could account for the lack of significance between groups as participants had little incentive to act dishonorably and thus, left little room for the commitment statement to truly have an effect. Therefore, we suggest future research focuses on the concept of binding commitment statements (i.e., commitments with punishment) that utilize a larger incentive scheme to see if under situations where the expected value high enough for loss aversion can take effect the results remain null or become significant. Lastly, the external validity of our research is also limited. One of the main functions of a commitment or contract is to "facilitate production and exchange over time" (Ederer & Schneider, 2019). In our experiment, second movers were asked to deliver on their commitment to trustworthy behavior immediately after agreeing to it. In contrast, making commitments and signing contracts in real-world scenarios often involves a delay between making the commitment and following through on the commitment. Future research should investigate the role of time in establishing trust between two individuals through pre-commitment mechanisms.

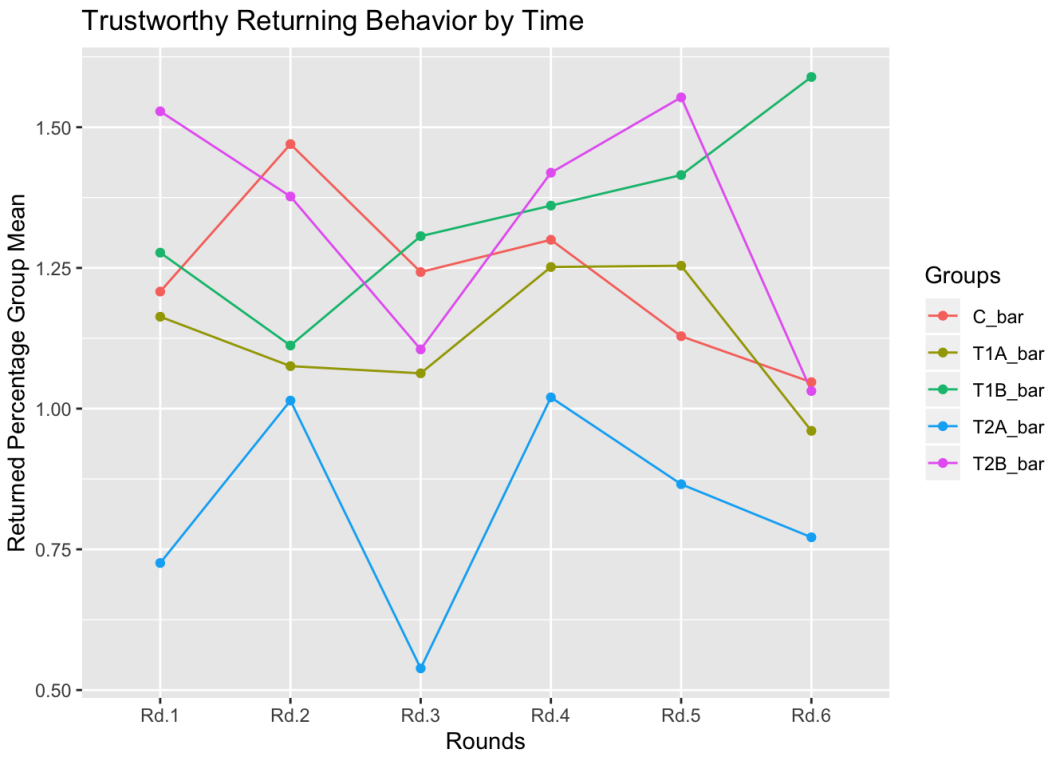


Figure 6. Trustworthy Returning Behavior by Time.

Challenges and Reflections

This research was conducted as part of a graduate-level course assignment at the University of Pennsylvania. As such, we were limited in our time and resources for running the experiment. Given more time and resources, we would have liked to design our own trust game on Qualtrics and add appropriate defaults to it, which would have made it easier for us as experimenters to run the study and easier for the participants to understand the game, perceive and react to the given instructions, and further proceed with the game. On the back end, this would have made our data analysis easier as well.

With our current study, the biggest challenge we faced was to make the participants understand the game and secondly make them comply to the given instructions. This was especially the case with Treatment 2, where a lot of our participants skipped agreeing to the mandatory commitment statement before logging in and starting the first set of games. This lack of compliance forced us to exclude several data points and further extended our data collection and recruitment time and effort. We encountered another significant challenge in regards to recruitment and sample population. We often recruited students who were in physical proximity to each other. Thus, during a particular data collection shift, several dyads were recruited from the same location on campus. This raises a particular concern since it is possible to assume that some of these participants share

very similar characteristics (i.e., all MBA students at Wharton playing against each other when recruited at Huntsman Hall), which may affect the external validity of our design. To overcome this challenge, we would ideally be recruiting participants from different locations on campus to play against each other.

CONCLUSIONS

Trust is important for human behavior, business settings, and public policy. Lawyers and business partners use trust and commitment contracts to establish trustworthiness to enhance engagement during business deals, and policy-makers can change collective behavior by improving commitment and trust among individuals. The goal of this study was to add to the existing body of literature by studying the role of pre-commitment and mid-commitment devices when building trustworthy behavior in game-playing conditions. Specifically, we seek to identify the direct effects commitments have on an individual's frequency of trustworthy behavior in the contexts of required and optional commitment statements. The study presented in this paper offers preliminary evidence that there is a preference for the current state of affairs, or a status quo bias, among students at the University of Pennsylvania. Furthermore, evidence shows that the presentation or signing of a commitment statement does not necessarily induce the desired behavior change, that is doesn't result in a greater trustworthy behavior between the two players. Future research should expand our research questions to investigate the causes and effects of this behavior further. Lastly, future research should investigate the role of the status quo bias in using commitment devices to build trustworthy behavior.

REFERENCES

Andreoni, J., & Serra-Garcia, M. (2019). The pledging puzzle: How can revocable promises increase charitable giving. CESifo Working Paper Series 7965, CESifo Group Munich.

Arrow, K. (1972). Gifts and exchanges. *Philosophy and Public Affairs, I*, 343–362.

Baca-Motes, K., Brown, A., Gneezy, A., Keenan, E. A., & Nelson, L. D. (2012). Commitment and behavior change: Evidence from the field. *Journal of Consumer Research, 39*(5), 1070-1084.

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

Berscheid, E., Boye, D., & Darley, J. M. (1968). Effect of forced association upon voluntary choice to associate. *Journal of Personality and Social Psychology, 8*(1p1), 13-19.

Bracht, J., & Feltovich, N. (2008). Efficiency in the trust game: An experimental study of precommitment. *International Journal of Game Theory, 37*(1), 39-72.

Breman, A. (2011). Give more tomorrow: Two field experiments on altruism and intertemporal choice. *Journal of Public Economics, 95*(11), 1349-1357.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: University Press, Princeton.

Ederer, F., & Schneider, F. (2019). The Persistent Power of Promises. *Available at SSRN 3169881*.

Fukuyama, F. (1995). *Trust*. New York: Free Press.

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology, 32*(5), 865-889.

Kellner, C., Reinstein, D., & Riener, G. (2019). Ex-ante commitments to "give if you win" exceed donations after a win. *The Journal of Public Economics, 169*, 109.

Levine, E. E., Bitterly, T. B., Cohen, T. R., & Schweitzer, M. E. (2018). Who is trustworthy? predicting trustworthy intentions and behavior. *Journal of Personality and Social Psychology, 115*(3), 468-494.

Lewis, D. A., Goetz, E., Schoenfield, M., Gordon, A. C., & Griffin, E. (1984). The negotiation of involuntary civil commitment. *Law & Society Review, 18*(4), 629-649.

Malhotra, D., & Murnighan, J. K. (2002). The effects of contracts on interpersonal trust. *Administrative Science Quarterly, 47*(3), 534-559.

Rogers, T., Milkman, K. L., & Volpp, K. G. (2014). Commitment devices: Using initiatives to change behavior. *Jama, 311*(20), 2065-2066.

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty, 1*(1), 7-59.

Schweitzer, M. E., Ho, T., & Zhang, X. (2018). How monitoring influences trust: A tale of two faces. *Management Science, 64*(1), 253-270.

Simpson, B., & Eriksson, K. (2009). The dynamics of contracts and generalized trustworthiness. *Rationality and Society, 21*(1), 59-80.

Wilkinson-Ryan, T. (2012). Transferring Trust: Reciprocity Norms and Assignment of Contract. *Journal of Empirical Legal Studies, 9*(3), 511-535.

APPENDIX A: EXPLORATORY DATA ANALYSES

Exploratory Analysis 1: Post-Hoc Power Analysis

Due to the lack of significance in our main results, we conducted a post-hoc power analysis to determine if the intervention truly had no effect, or there was a type II error (false negative). For the null effect in result 1, the effect size calculated in G*Power was around 0.255 and the corresponding power achieved was 16.14%. This calculation means that there is an 83.86% chance that the study actually has a significant result but we end up reporting the null effect. Additionally for result 2, an effect size of 0.36 reached a power of 27.45%. The value of both two post-hoc power was much lower than intended 80%, significantly increasing the possibility of type II error. Therefore, although the commitment statement in our experiment might not have any effect, the small sample size and power could affect the credibility of this insignificance.

Exploratory Analysis 2: Regression Analysis

After conducting a statistical test for each hypothesis, we generated the corresponding regression model to further confirm the result discussed in the main paper and distinguish demographic factors potentially changing the effect. Pilot data were not used in the regression analysis since information about age and experience of trust game was not collected.

As shown in the first column of ***Figure A1***, we regressed the incidence of trustworthy behavior in Game 1 on independent variables including agreeing to a commitment statement or not before Game 1, agreeing to a commitment statement or not before Game 2, age (18-31), gender, school level (6 categories in total), and having experience of trust game or not before the study within all second movers in Control, Treatment 1 and Treatment 2 groups. Consistent with the outcome of the Wilcoxon test discussed above, agreeing to a commitment statement did not contribute significantly to the decision of trustworthy returning behavior.

Column 2 summarizes the relationship between the midway point self-selection and the rest of the independent variables. 54 observations with second movers in Treatment 1 and Treatment 2 were included. The coefficient of 0.378 with a significant level of 1% for agreeing to a commitment statement before Game 1 again matches with our previous discussion. Agreeing to a mandatory initial commitment statement was positively related to the self-selection to agree to the new midway point commitment statement in the following game.

Columns 3 and 4 both represent the relationship between the incidence of trustworthy behavior

in Game 3 on independent variables the same as the first regression. Since H2 focused on comparisons between Treatment 1 and Treatment 2, while H3 tried to rank order across the groups including Control, the number of observations was the only difference in the construction of those two regressions. We can see the midway point selection with a coefficient of 0.245 was significant at a 10% significance level in column 3, showing some plausible effect on behavioral change with commitment statement. However, we still concluded a null effect for H2 before since we usually required a significance level of 5%.

It is interesting to see that among all regressions, demographic variation was not making any significant difference in trustworthy behavior and decision we studied. In fact, R-squared, the statistical measure of goodness of fit, was very small for all the four regression, either indicating a linear model was suspicious or some most important variables were overlooked.

Regression Table for Hypotheses

	Dependent variable:			
	Trustworthyincid_phase1	Commitment..1.yes...0.no..1	Trustworthyincid_phase2	
	(1)	(2)	(3)	(4)
Commitment..1.yes...0.no.	0.096	0.378***	0.118	0.044
	(0.125)	(0.136)	(0.122)	(0.121)
Commitment..1.yes...0.no..1	-0.092		0.245*	0.158
	(0.117)		(0.124)	(0.113)
Age	0.001	-0.010	-0.006	-0.009
	(0.022)	(0.031)	(0.026)	(0.021)
Gender..male.0..female.1..other.2..prefer.not.to.say.3.	-0.124	-0.079	-0.018	0.044
	(0.096)	(0.137)	(0.114)	(0.093)
School.LevelGrad-NOT Penn	0.188	0.336	-0.493	-0.403
	(0.488)	(0.574)	(0.478)	(0.470)
School.LevelGrad-Penn	0.022	0.122	-0.161	-0.088
	(0.229)	(0.300)	(0.250)	(0.220)
School.LevelNOT Student	-0.155	0.284	0.011	-0.065
	(0.385)	(0.595)	(0.495)	(0.371)
School.LevelUndergrad-NOT Penn	-0.384			-0.424
	(0.490)			(0.472)
School.LevelUndergrad-Penn	0.183	0.215	0.011	-0.087
	(0.264)	(0.350)	(0.291)	(0.254)
Trust.Game.Experience..yes.1..no.0.1	-0.094	0.042	-0.177	-0.084
	(0.101)	(0.145)	(0.121)	(0.097)
Constant	0.450	0.513	0.610	0.735
	(0.605)	(0.855)	(0.712)	(0.584)
Observations	84	54	54	84
R^2	0.088	0.195	0.212	0.087
Adjusted R^2	-0.037	0.052	0.051	-0.038
Residual Std. Error	0.425 (df = 73)	0.479 (df = 45)	0.398 (df = 44)	0.410 (df = 73)
F Statistic	0.706 (df = 10; 73)	1.360 (df = 8; 45)	1.315 (df = 9; 44)	0.698 (df = 10; 73)
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure A1. Regression table for each hypothesis. Regression (1), (2), (3), (4) stand for H1a, H1b, H2, and H3, respectively.

Exploratory Analysis 3: Retrospective Analysis

Through our analysis, there was no statistically significant difference in comparisons between

any two groups. However, we also wanted to understand what motivated participants to recommit to the commitment statement during the midway point of the game. Thus, initially, we treated all participants within Treatment 2 as a whole in Game 1, but now we retrospectively separated them according to their midway point self-selection. In other words, we hypothetically had T2A and T2B for Game 1 as well. As shown in ***Figure A2***, the density distribution of the incidence of trustworthy returning behavior was much more skewed to the right in the left segment. The p-value from a Wilcoxon test between those two groups was 0.0388, which means that in Game 1, participants who self-selected to re-commit later behaved in a more trustworthy way than those who did not. This result leads to a potential claim that people who were forced to agree to the commitment statement would neither be truly more trustworthy nor recommit subsequently, while the nature of trustworthiness might be the reason behind the voluntary re-commitment before Game 2.

	T2A (Phase 1)	T2B (Phase 1)
Dyads of participants	n=7	n=23
Incidence Group Mean	0.190 (0.178)	0.551(0.397)

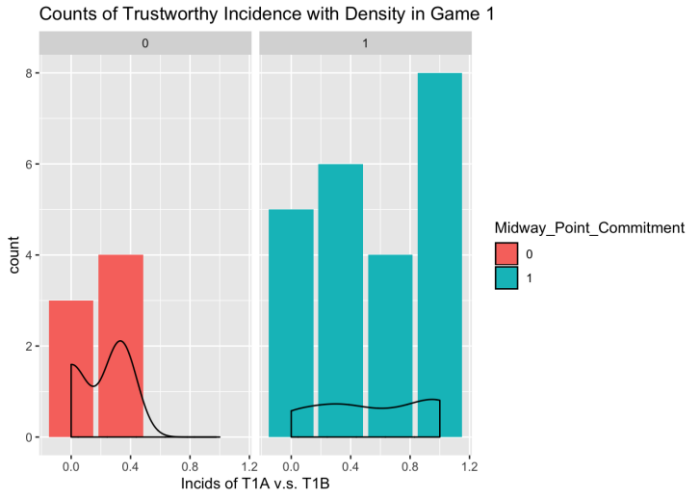


Figure A2. Barplots of the counts of retrospective trustworthy incidence with density distribution in Game 1. Orange bars represent second movers who did not agree to the midway commitment statement, and blue bars represent those who did agree.

APPENDIX B: ONLINE GAMEPLAY INTERFACE

Participants played standard 2-player trust games powered by *oTree* through the website economic-games.com (see ***Figure A3*** and ***Figure A4***). During gameplay, experimenters logged into the game interface as administrators to follow and analyze each move (see ***Figure A5***).

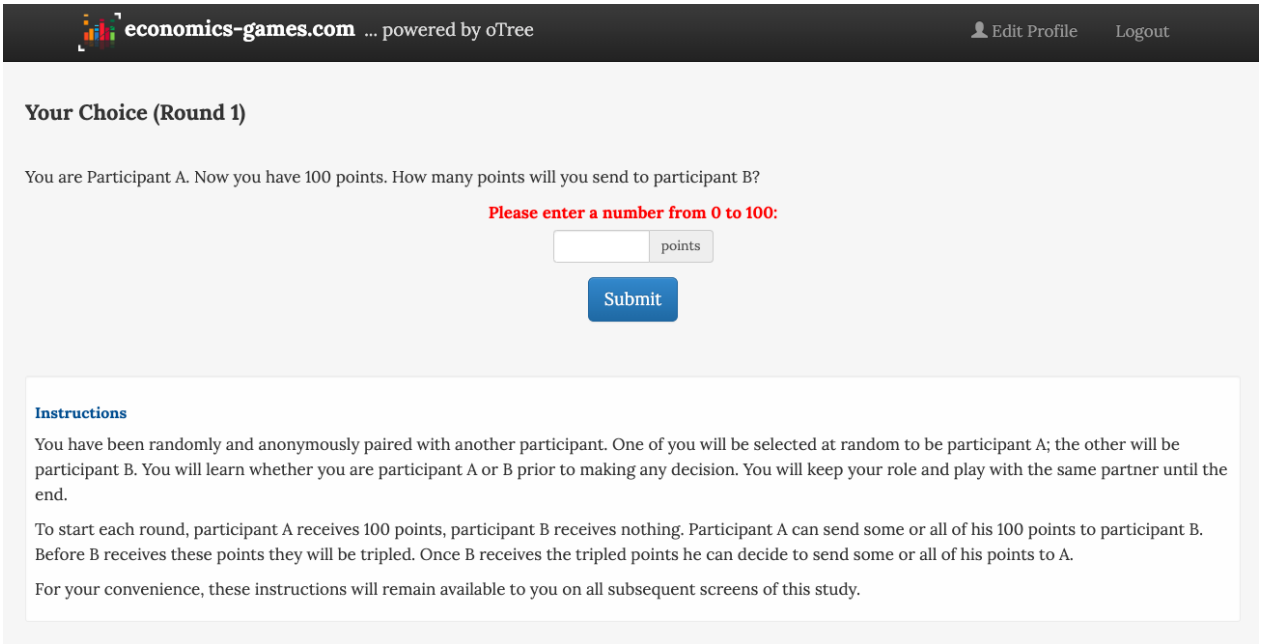


Figure A3. *Screenshot of Player A's first move decision from economic-games.com.* Player A must decide how many of their initial 100 points to send to Player B, knowing that these points will be tripled for Player B.

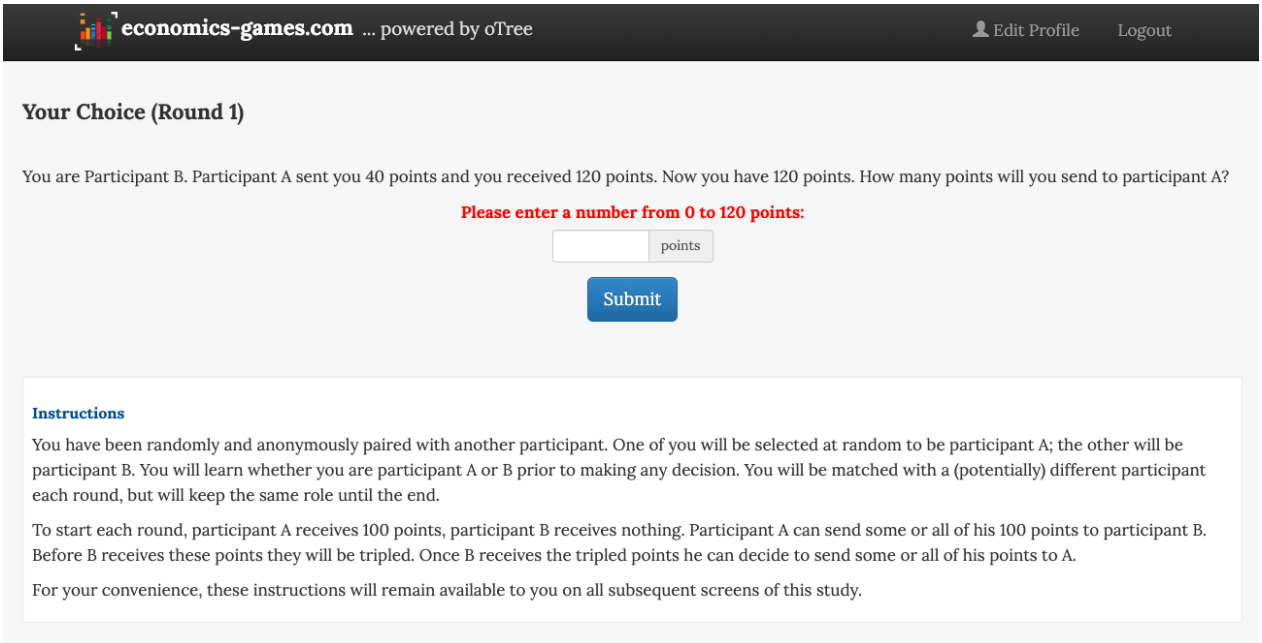
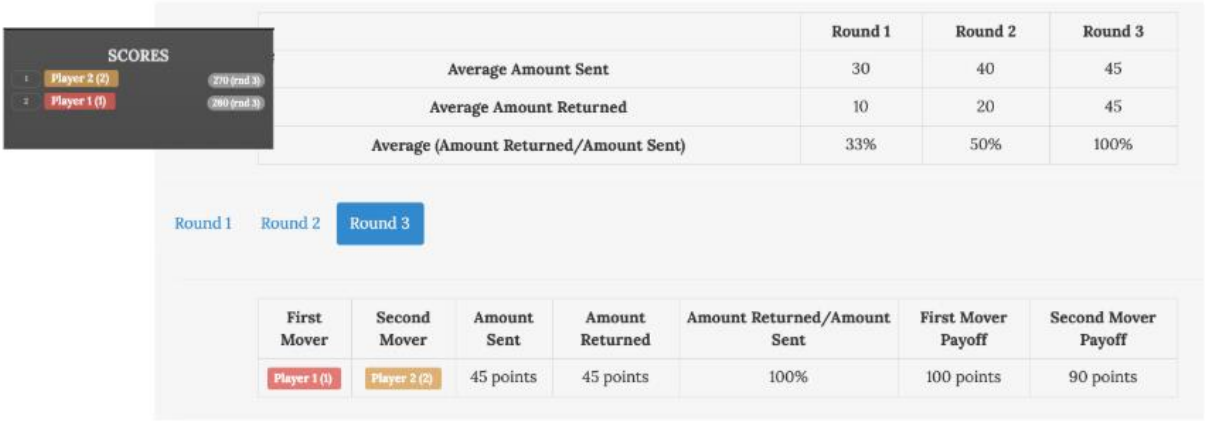


Figure A4. *Screenshot of Player B's first move decision from economic-games.com.* Player B must decide how many of the now tripled points they received from Player A to send back to Player A.



	Round 1	Round 2	Round 3
Average Amount Sent	30	40	45
Average Amount Returned	10	20	45
Average (Amount Returned/Amount Sent)	33%	50%	100%

First Mover	Second Mover	Amount Sent	Amount Returned	Amount Returned/Amount Sent	First Mover Payoff	Second Mover Payoff
Player 1 (1)	Player 2 (2)	45 points	45 points	100%	100 points	90 points

Figure A5. *Screenshot of Administrator Summary View from economic-games.com.* Experimenters logged into the game interface as administrators could track players' moves in real-time and collect summary statistics for each round.

APPENDIX C: PARTICIPANT INSTRUCTIONS

Exhibit 1. *Envelope 1 Instructions for First Movers (All Conditions)*

Welcome! In this study, you will have the opportunity to earn raffle tickets for a $100 Amazon Gift Card by participating in economic interactions with a partner. Regardless of the outcome, you and your partner will each be given 1 raffle ticket for participation and have the opportunity to earn more raffle tickets based on you and your partner's decisions during the interactions. The number of raffle tickets you will receive is proportional to the number of points you have at the end of the game.

We will ask you to play an economic game on the computer screen in front of you. Everything you need will be in the envelopes in front of you. Please open the envelopes in numerical order and when instructed. If you have any questions during the experiment, please raise your hand and the experimenter will come in.

To begin, follow these steps:

1. Click "**Login**" on the screen (upper right-hand corner)

2. Enter the "**Login**" and "**Password**" below
   - Login: _____
   - Password: **pass**
3. Select "**Proceed to Game**"
4. Do not change your team name, simply select "**Submit**"
5. Read the instructions for the game carefully. (You will be asked an understanding question before you begin the online games)

When the game is over, you will see a screen stating:
The game is over, thank you for playing!

Please select "**Logout**" in the upper right-hand corner

At this time, you can **open Envelope #2**.

Exhibit 2. *Envelope 2 Instructions for First Movers (All Conditions)*

Good job! You have completed the first half of the study. For the second half, you will play the same type of games following exactly the same steps you just completed. You will be interacting with the **same partner**.

To begin the second half, follow these steps:

1. Click "**Login**" on the screen (upper right-hand corner)

2. Enter the "**Login**" and "**Password**" below
   - Login: _____
   - Password: **pass**
3. Select "**Proceed to Game**"
4. Do not change your team name, simply select "**Submit**"
5. Read the instructions for the game carefully. (You will be asked an understanding question before you begin the online games)

When the game is over, you will see a screen stating:
The game is over, thank you for playing!

At this time, you can **open Envelope #3**.

Exhibit 3. *Envelope 3 Instructions for First Movers (All Conditions)*

Thank you for participating!

Before you leave, please respond to the following questions:

1. What is your age? _____

2. What is your gender?

   - ❑ Female
   - ❑ Male
   - ❑ Other
   - ❑ Prefer not to say

3. What is your current level of schooling?

   - ❑ Undergrad Student at Penn
   - ❑ Grad Student at Penn
   - ❑ Doctoral Student at Penn
   - ❑ Undergrad Student not at Penn
   - ❑ Grad Student not at Penn
   - ❑ Doctoral Student not at Penn
   - ❑ Not currently a student

4. Have you learned about or played an economic trust game like the ones in this experiment before?
   - ❑ Yes
   - ❑ No

5. Please provide your email address so we can reach out if you win the $100 Amazon Gift Card raffle:

   _____

The experiment is now complete. Please leave all materials inside the room. When you exit, the experimenter will inform you of the total number of raffle tickets you earned during the online interactions in addition to the initial raffle ticket you receive for your time and participation.

Exhibit 4. *Envelope 1 Instructions for Second Movers (Control and Treatment 1)*

Welcome! In this study, you will have the opportunity to earn raffle tickets for a $100 Amazon Gift Card by participating in economic interactions with a partner. Regardless of the outcome, you and your partner will each be given 1 raffle ticket for participation and have the opportunity to earn more raffle tickets based on you and your partner's decisions during the interactions. The number of raffle tickets you will receive is proportional to the number of points you have at the end of the game.

We will ask you to play an economic game on the computer screen in front of you. Everything you need will be in the envelopes in front of you. Please open the envelopes in numerical order and when instructed. If you have any questions during the experiment, please raise your hand and the experimenter will come in.

To begin, follow these steps:

1. Click "**Login**" on the screen (upper right-hand corner)

2. Enter the "**Login**" and "**Password**" below
   - Login: _____
   - Password: **pass**
3. Select "**Proceed to Game**"
4. Do not change your team name, simply select "**Submit**"
5. Read the instructions for the game carefully. (You will be asked an understanding question before you begin the online games)b

When the game is over, you will see a screen stating:
The game is over, thank you for playing!

Please select "**Logout**" in the upper right-hand corner

At this time, you can **open Envelope #2**.

Exhibit 5. *Envelope 1 Instructions for Second Movers (Treatment 2)*

For T2 second movers, Envelope 1 instructions were modified to include a *mandatory*

commitment statement on the first page.

Welcome! In this study, you will have the opportunity to earn raffle tickets for a $100 Amazon Gift Card by participating in economic interactions with a partner. Regardless of the outcome, you and your partner will each be given 1 raffle ticket for participation and have the opportunity to earn more raffle tickets based on you and your partner's decisions during the interactions. The number of raffle tickets you will receive is proportional to the number of points you have at the end of the game.

We will ask you to play an economic game on the computer screen in front of you. Everything you need will be in the envelopes in front of you. Please open the envelopes in numerical order and when instructed. If you have any questions during the experiment, please raise your hand and the experimenter will come in.

Before you begin, ***you must agree*** to the following statement committing to behaving in a trustworthy manner. (For the purposes of this research, we define "trustworthy" behavior as actions one party takes that fulfill both their own interests and the interests of the other party.)

To commit, please rewrite the following statement and sign your initials below it.

"I agree to act trustworthy in the following interactions."

Commitment Statement: _____
_____

Initials: _____

PLEASE SEE REVERSE FOR NEXT STEPS

To begin, follow these steps:

1. Click "**Login**" on the screen (upper right-hand corner)

2. Enter the "**Login**" and "**Password**" below
   - Login: _____
   - Password: **pass**
3. Select "**Proceed to Game**"
4. Do not change your team name, simply select "**Submit**"
5. Read the instructions for the game carefully. (You will be asked an understanding question before you begin the online games)

When the game is over, you will see a screen stating:
The game is over, thank you for playing!

Please select "**Logout**" in the upper right-hand corner

At this time, you can **open Envelope #2**.

Exhibit 6. *Envelope 2 Instructions for Second Movers (Control)*

Good job! You have completed the first half of the study. For the second half, you will play the same type of games following exactly the same steps you just completed. You will be interacting with the **same partner**.

To begin the second half, follow these steps:

1. Click "**Login**" on the screen (upper right-hand corner)

2. Enter the "**Login**" and "**Password**" below
   - Login: _____
   - Password: **pass**
3. Select "**Proceed to Game**"
4. Do not change your team name, simply select "**Submit**"
5. Read the instructions for the game carefully. (You will be asked an understanding question before you begin the online games)

When the game is over, you will see a screen stating:
The game is over, thank you for playing!

At this time, you can **open Envelope #3**.

Exhibit 7. *Envelope 2 Instructions for Second Movers (Treatment 1)*

For T1 second movers, Envelope 2 instructions were modified to include a *voluntary*

commitment statement on the first page.

Good job! You have completed the first half of the study. For the second half, you will play the same type of games following exactly the same steps you just completed. You will be interacting with the **same partner**.

Before you continue, please carefully consider the following commitment to your partner. You may *choose* to agree to the following statement committing to behaving in a trustworthy manner. (For the purposes of this research, we define "trustworthy" behavior as actions one party takes that fulfill both their own interests and the interests of the other party.)

Committing to this statement is **completely voluntary.**

If you choose not to commit, you may leave this part blank.

If you choose to commit, please rewrite the following statement and sign your initials below it.

"I agree to act trustworthy in the following interactions."

Commitment Statement: _____
_____

Initials: _____

PLEASE SEE REVERSE FOR NEXT STEPS

To begin the second half, follow these steps:

1. Click "**Login**" on the screen (upper right-hand corner)

2. Enter the "**Login**" and "**Password**" below
   - Login: _____
   - Password: **pass**
3. Select "**Proceed to Game**"
4. Do not change your team name, simply select "**Submit**"
5. Read the instructions for the game carefully. (You will be asked an understanding question before you begin the online games)

When the game is over, you will see a screen stating:
The game is over, thank you for playing!

At this time, you can **open Envelope #3**.

Exhibit 8. *Envelope 2 Instructions for Second Movers (Treatment 2)*

For T2 second movers, Envelope 2 instructions were modified to include a *voluntary* re-commitment statement on the first page.

Good job! You have completed the first half of the study. For the second half, you will play the same type of games following exactly the same steps you just completed. You will be interacting with the **same partner**.

Before you continue, please carefully consider the following additional commitment to your partner. You may *choose* to agree to the following statement re-committing to behaving in a trustworthy manner. (For the purposes of this research, we define "trustworthy" behavior as actions one party takes that fulfill both their own interests and the interests of the other party.)

Re-committing to this new statement is **completely voluntary.**

If you choose not to re-commit, you may leave this part blank.

If you choose to re-commit, please rewrite the following statement and sign your initials below it.

"I agree to act trustworthy in the following interactions."

Commitment Statement: _____
_____

Initials: _____

PLEASE SEE REVERSE FOR NEXT STEPS

To begin the second half, follow these steps:

1. Click "**Login**" on the screen (upper right-hand corner)

2. Enter the "**Login**" and "**Password**" below
   - Login: _____
   - Password: **pass**
3. Select "**Proceed to Game**"
4. Do not change your team name, simply select "**Submit**"
5. Read the instructions for the game carefully. (You will be asked an understanding question before you begin the online games)

When the game is over, you will see a screen stating:
The game is over, thank you for playing!

At this time, you can **open Envelope #3**.

Exhibit 9. *Envelope 3 Instructions for Second Movers (All Conditions)*

Thank you for participating!

Before you leave, please respond to the following questions:

1. What is your age? _____

2. What is your gender?

   - ❑ Female
   - ❑ Male
   - ❑ Other
   - ❑ Prefer not to say

3. What is your current level of schooling?

   - ❑ Undergrad Student at Penn
   - ❑ Grad Student at Penn
   - ❑ Doctoral Student at Penn
   - ❑ Undergrad Student not at Penn
   - ❑ Grad Student not at Penn
   - ❑ Doctoral Student not at Penn
   - ❑ Not currently a student

4. Have you learned about or played an economic trust game like the ones in this experiment before?
   - ❑ Yes
   - ❑ No

5. Please provide your email address so we can reach out if you win the $100 Amazon Gift Card raffle:

   _____

The experiment is now complete. Please leave all materials inside the room. When you exit, the experimenter will inform you of the total number of raffle tickets you earned during the online interactions in addition to the initial raffle ticket you receive for your time and participation.

APPENDIX D: EXPERIMENTAL SETUP

The experiment was set up simultaneously at Van Pelt Library, Huntsman Hall, and Sansom Place

East (meeting rooms of the graduate student residence hall). The pictures below depict our researchers observing an on-going trust game between two players.