INTERDISCIPLINARY CENTER
FOR ECONOMIC SCIENCE
GEORGE MASON UNIVERSITY

Harnessing the Benefits of Betrayal Aversion

Jason A. Aimone and Daniel Houser

January 2012

Discussion Paper

Interdisciplinary Center for Economic Science
4400 University Drive, MSN 1B2, Fairfax, VA 22030
Tel: +1-703-993-4719   Fax: +1-703-993-4851
ICES Website: http://ices.gmu.edu
ICES RePEc Archive Online at: http://edirc.repec.org/data/icgmuus.html

Harnessing the Benefits of Betrayal Aversion

by

Jason A. Aimone[*]

And

Daniel Houser[!]

[*]Virginia Tech Carilion Research Institute

[!]Interdisciplinary Center for Economic Science

George Mason University

August 2011

Abstract:
Recent research suggests that while there are negative effects of betrayal aversion, that the presence of betrayal-averse agents is beneficial in reducing trustees' willingness to betray trust. If true, then many common knowledge institutions may have adopted institutional rules and features which mitigate the emotional disutility associated with betrayal aversion while simultaneously maintaining the high levels of reciprocation brought about by the presence of betrayal-averse agents. Here we conduct a laboratory experiment which identifies a prevalent successful institutional feature common to many every-day institutions: the voluntary, but not forced, option to discover the painful details of failed economic exchange.

JEL Classifications: *C91; D03; D81*

Keywords: Betrayal Aversion, Risk, Trust, Institutions

Acknowledgements: For helpful comments and suggestions we thank Omar Al-Ubaydli, Vernon Smith, Virgil Storr, Bart Wilson, and seminar participants at George Mason University, The University of Kansas, Villanova University and The Virginia Tech Carilion Research Institute. The authors gratefully acknowledge financial research support from the International Foundation for Research in Experimental Economics, The Mercatus Center, and the National Science Foundation Research Grant SES-0851250. Adam Smith and Jared Barton provided excellent laboratory assistance.

1 Introduction

Recent experimental work (e.g. Bohnet and Zeckhauser, 2004; and Aimone and Houser, 2009) suggests that the knowledge of a personal betrayal of trust creates a disutility that cannot be fully explained by pecuniary losses. Indeed, a significant portion of the distrust we observe in the lab may stem from the desire to avoid the feeling of being betrayed, or "betrayal aversion" (this might be considered a specific type of emotion regulation strategy, see, e.g., Gross, 1998; Ochsner and Gross, 2005; Mui et al., 2008). Nevertheless, follow-up work by Aimone and Houser (2011, henceforth AH2011) shows that the presence of betrayal-averse agents actually generates reciprocity, thereby stimulating trust. This dual nature of betrayal aversion poses a dilemma for institution design. Specifically, betrayal-averse agents are less apt to trust than non-betrayal averse agents, leaving social gains unrealized. Yet the institutional elimination of betrayal aversion creates its own inefficiencies, namely increased rates of betrayal (AH2011). How, then, can we design institutions with these conflicting effects of betrayal aversion in mind? We demonstrate experimentally that it is possible to create an institution that maintains relatively low rates of betrayal while stimulating significantly increased rates of trust. To do so, we do not eliminate the possibility of active betrayal-averse agents, but rather make the knowledge of betrayal optional.

Many previously-studied institutions, such as punishment, repeat play, and the addition of reputation effects, indirectly reduce the effect of betrayal aversion by reducing the *probability* that a betrayal-averse agent will experience the emotional disutility associated with betrayal. Nevertheless, these often-studied institutions do not reduce the *magnitude* of the emotional disutility associated with a given betrayal. Thus, they leave gains unrealized, as betrayal-averse agents remain relatively less likely to trust than non-betrayal-averse agents.

Firms regularly adopt institutions that allow people to avoid the knowledge of a personal betrayal. These institutions, which remain relatively unexplored, work to reduce the magnitude of disutility associated with betrayal aversion (or eliminate it all together). For example, consider a chain restaurant. Upon receiving bad food, a customer has several options, which range from doing nothing at all to complaining to the waiter, the cook, the manager, or the corporate offices. Doing nothing shields the customer from potentially discovering that the betrayal was personal in nature, while an inquiry or complaint could reveal that the low-quality product (the bad food) was the result of a personal betrayal, e.g., a waiter or cook serving food that they dropped on the floor, or the company purchasing sub-standard ingredients.[1] We suggest, with support from a generalized version of this type of institution in the experimental study, that institutions with optional betrayal knowledge capture the beneficial aspects of betrayal aversion while mitigating the negative aspects.

Previous research has identified betrayal aversion as a factor that both detrimentally and beneficially affects social economic exchange. Betrayal aversion leads us to trust others less frequently (Bohnet and Zeckhauser, 2004: Bohnet et al., 2008, Aimone and Houser 2009), punish betrayers more severely (Koehler and Gershoff, 2003), and choose more hazardous safety products (Koehler and Gershoff, 2011). These negative effects are offset, however, by the increased rates of reciprocation trustees show toward potentially betrayal-averse trustors (Aimone and Houser, 2011); likewise, they may also be mitigated by a natural neuropeptide (oxytocin) that increases our willingness to trust when around intimate others (Kosfeld et al., 2005).

A few recent studies indirectly provide some insight into the institutional side of betrayal aversion. Birnberg and Zhang (2010) show that the expectation of betrayal by agents leads principals to choose inefficient institutions that remove the possibility of betrayal. In Bohnet et al. (2010), the authors show evidence that differences in elasticities of trust between Middle Eastern and western countries can be traced in part to differences in how their institutions handle potential acts of betrayal. Nevertheless, these studies do not attempt to disentangle betrayal aversion from other motivating factors, like desire for accountability, loss-aversion or altruism. Further, none of the previous work on betrayal aversion or institutions provides guidance or evidence on how institutions have been, or could be, designed to harness the beneficial effects of betrayal aversion and mitigate the negative effects. Thus, there remains a gap between the exploratory laboratory analysis of betrayal aversion and future empirical applied studies of institutions and betrayal aversion.

Our laboratory experiment bridges this gap by presenting and studying a generalized common knowledge institution found in many everyday institutions. We identify the institution as one that successfully mitigates negative effects of betrayal aversion without suffering a

[1] Similarly, the addition of refunds and warranties are additional institutional features that allow customers to recoup losses without ever being exposed to the specific betrayer, the person who determined the product's low quality. This leaves the act of betrayal impersonal.

detrimental increase in rates of betrayal. The proposed institution makes the knowledge of betrayal optional for investors, without forcing them to know whether their specific counterpart trustee chose to betray their trust. We implement this institution in the laboratory using a standard binary version of the trust game (Berg et al. 1995).

Our results demonstrate that the institution successfully achieves significantly increased rates of trust. Our key finding is that this increased frequency of trust occurs without the adverse effect of increased rates of betrayal (previously observed in Aimone and Houser 2011). This result suggests that the possibility that one's counterpart could be exposed to negative emotions associated with betrayal aversion is sufficient for the institution to maintain the beneficial aspects of betrayal aversion.

The following section presents a detailed background on how previous studies have demonstrated that betrayal aversion both negatively influences trust and positively stimulates trust and reciprocity. Section 3 explains our laboratory design and hypotheses. Section 4 presents our results. Section 5 concludes by examining how many common institutions successfully remove the negative aspects of betrayal aversion.

2 Background on Betrayal Aversion

Many different fields of research study betrayal aversion, focusing predominately upon identifying the existence and negative effects of betrayal aversion. Generally, these studies isolate betrayal aversion by comparing the behavior of agents under institutions that either expose the trusting agent to betrayal knowledge or fully remove betrayal knowledge.

In Koehler and Gershoff (2003 and 2011), the authors used paid surveys to observe how responses to similar hypothetical situations differ based on how the situation is framed. Their survey subjects reported that crimes involving acts of betrayal (such as a security guard committing robbery) should correspond to more years in prison than crimes without a trust betrayal (such as a janitor committing robbery). The authors interpreted this penalty premium as betrayal aversion. Likewise, they showed that this form of betrayal aversion is not only related to human betrayals; indeed, people are also averse to betrayal from inanimate objects of safety, such as vaccines or airbags. Survey respondents in their studies reported that they would rather have a less effective safety device, with an overall higher probability of death, than a more effective safety device with a minute probability of harming the owner on its own.

In another experiment, Birnberg and Zhang (2010) studied accountability and betrayal using a principal and agent environment where principals and agents shared reported earnings 50/50. In the study, only agents were able to observe the actual earnings of the subject pair (if not prevented by the institution). Therefore, agents had the option to falsely report low earnings, even if the pair had actually achieved high earnings. In this way, they were effectively able to "embezzle" earnings from the principal. Birrnberg and Zhang demonstrated that a large fraction of principals would prefer a "restrictive" institution to a "permissive" institution. The restrictive institution would completely prevent agent embezzlement, but reduce earnings to a level which would be reached only if 100% of agents chose to embezzle. Given that principals reported expecting less than 100% of agents to embezzle earnings, and that less than 100% actually choose to steal, the authors attributed the significant portion of principals still choosing the restrictive institution as evidence that those principals either had a "desire for accountability" or were betrayal-averse.

In Bohnet and Zeckhauser (2004), Hong and Bohnet (2007) and Bohnet et al. (2008), experimenters compared decisions in trust and risky dictator "MAP" games. In these games, a first mover reported the "minimum acceptable probability" that they would be willing to either trust a counterpart (in the trust version of the game) or engage both themselves and a counterpart in a lottery (in the risky dictator version of the game). If the first mover's reported MAP was less than a predetermined percentage, 'p*', the first and second mover were paid based upon the result of the respective game. If the reported MAP was larger than p*, the agent did not enter the trust gamble; he/she received $10 and his/her counterpart received $10. If the trust game was played, the first mover as paid according to his/her own second mover counterpart's decision to reciprocate the trust (in which case $30 was split evenly) or betray the trust (in which case the second mover kept $22 and the first mover received only $8.) If the risky dictator game was played, both agents were paid based upon the outcome of a risky lottery, where the two monetary outcomes of the lottery were the same as the outcomes of the binary trust game (first mover (second mover) get $15($15) or $22($8)). In the trust game, p* was determined by the true percentage of second movers in that session that chose to reciprocate trust. In the risky dictator game, p* equaled the average p* observed in the trust game sessions. Studies using this design attribute the higher MAPs reported in the trust game (as compared to the risky dictator game) to betrayal aversion.[2]

Distinguishing the different effects of institutions is particularly relevant when performing cross-country comparisons of institutions. Bohnet et al. (2010) demonstrated that the elasticity of trust differs greatly in Western and Middle Eastern countries. They attributed the difference to the adoption, in Western Cultures, of relatively more institutions focused on the mitigation of losses from betrayal (such as insurance). In contrast, Middle Eastern cultures tend to adopt relatively more institutions designed to prevent betrayals (such as threat of expulsion from the group in the event of a betrayal.) Indeed, their subjects in the Middle East had greater expectations of reciprocations than their subjects from the west. As a result, the Middle Eastern subjects exhibited higher reference points of trust and lower elasticity of trust relative to the subjects from the west

Nevertheless, as previously discussed, a "prevention" regime, like those found in the Middle East, may reduce the probability of betrayal while increasing the disutility associated with the experience of betrayal. This increased betrayal aversion from a prevention regime could manifest itself in an overall decreased frequency of trust and participation in social exchange relative to "mitigation" regimes, where reference-points of trust are lower, and disutility associated with betrayal aversion is reduced. It is possible for both mitigation and prevention regimes to incorporate an institution that makes the knowledge of personal betrayal optional for investors. We demonstrate that this type of institution can both mitigate the disutility associated with betrayal aversion and prevent increased rates of betrayal, resulting in an overall significantly increased frequency of trust.

In this study, we adopt the betrayal aversion identification procedure used by Aimone and Houser (2009 & 2011), henceforth AH2009 and AH2011. Therefore, we spend more time explaining the details of their design. Aimone and Houser pointed out that in a standard trust game (e.g., AH2009's baseline "KNOW" treatment seen in Figure 1), when an investor chooses to trust, the amount of money he/she earns reveals two pieces of information: i) the monetary outcome of exchange (either both investor and trustee receive $15 or the investor receives $2 and the trustee receives $28); and ii) whether his/her counterpart trustee reciprocated or betrayed

[2] Bohnet et al. 2010 interpret this difference as reflecting loss aversion.

trust. Their "DONTKNOW" treatment removes this second piece of information by basing some investors' pay on random draws from the pool of trustee decisions (reciprocate ($15) or betray ($2)) rather than their own trustee's decision.[3] Thus, a payment of $2 ($15) only reveals that *some* trustee betrayed (reciprocated) trust, and not whether the investor's *own counterpart* trustee (the trustee whose payoff is affected by the investor's decision to trust) betrayed or reciprocated trust.

<Figure 1>

Aimone and Houser found that significantly more investors trust when the knowledge of a personal betrayal is removed from the environment (from 65.4% of investors in KNOW to 92% in DONTKNOW) (p<0.03, Mann-Whitney two-tailed). Similarly, in their "OPTION" treatment, when they gave some investors a choice on whether to be paid based upon their own counterpart's betrayal decision or the randomly drawn decision, they found that all investors were willing to risk the $5 they would get from the safe option in order to take the trust gamble (54% choose to avoid the knowledge of personal betrayal or reciprocation). These results indicate that significant portions of investors are averse to the knowledge of personal betrayal, independent of the monetary outcomes (and risk) of the trust environment.

While trustees in AH2009 were not informed about the institutional feature which removed betrayal aversion for some subjects[4], in AH2011 the authors made the institutional framework in DONTKNOW common knowledge to both investors and trustees in a new treatment, "DONTKNOW2." They found that without the influence of betrayal aversion, trustee betrayal rates increased significantly (from 66.3% in the "BASELINE" treatments (KNOW, DONTKNOW, and OPTION) to 85% in DONTKNOW2 (p<0.04)). Similarly, trust rates dropped from the 92.0% in DONTKNOW to 58.8% in DONTKNOW2 (p<0.01) when the institution that removed betrayal aversion became common knowledge. The authors interpreted these results as the first evidence of the beneficial aspects of betrayal aversion. They suggested that the presence of betrayal-averse agents increases rates of reciprocation. They attributed this

[3] Their own trustee receives payment based upon their own decision to betray ($28) or reciprocate ($15) trust, and is not aware that some investors were able to avoid the knowledge of specific betrayal.

[4] Note that there was no deception used in AH2009 or AH2011, see their papers for details. We similarly adopt a strict no deception rule.

increase to trustee's other-regarding preferences, and suggested that it would lead to more trust by investors due to the fact that the monetary environment is less risky.

In summary, these past studies have: i) identified the presence of preferences consistent with betrayal aversion; ii) determined that betrayal aversion can both negatively and positively affect economic behavior; iii) found evidence consistent with the widespread existence of betrayal aversion, which varies across cultures and countries; and iv) found evidence suggesting economic agents are averse to betrayal both from animate and inanimate objects. Additionally, Koheler and Gershoff (2011) provide some evidence that betrayal aversion to inanimate object betrayal is responsive to framing effects in a hypothetical decision environment.[5] Researchers have not yet begun to analyze how agents respond to social exchange institutions that affect betrayal aversion without completely removing betrayal aversion or leaving it unchanged. While the outcomes of OPTION treatment (in AH2009) are appealing (100% investment in profitable trust with low levels of betrayal), the institution could not be implemented in naturally occurring environments as the trustee side of the market is unaware of the institution. As such, we are not aware of any studies that investigate the rules within naturally-occurring institutions that accommodate betrayal-averse agents. Here, we present first evidence on the responsiveness of agents to a common-knowledge institution designed to reduce the negative aspects of betrayal aversion while harnessing the benefits of betrayal aversion.

3 Design and Hypotheses

3.1 Design Motivation

We seek to test an institution that possesses the following four features: i) investors can avoid the knowledge of personal betrayal; ii) investors face the same probability of a reciprocation or betrayal of their trust regardless of whether they are exposed to the knowledge of personal betrayal; iii) trustees know that investors have the option to avoid the knowledge of personal betrayal; iv) trustees make the decision to betray or reciprocate trust without knowing whether the investor will be exposed to the knowledge of personal betrayal.

Outside the laboratory, many existing economic environments contain all of these features; however, these environments also normally contain additional institutional features that

[5] Betrayal aversion is reduced if the betrayal is presented as passive instead of active, if positive emotions are primed, or if analytical instead of emotional decision-making is primed.

would confound attempts to empirically study the effects of betrayal aversion within the environments. Such factors could include communication, promises, reprisal, retribution, the possibility of punishment (either pecuniary or non-pecuniary), and many others. The institution would thus be expected to interact with these additional factors, preventing researchers from attributing outcomes to betrayal aversion or the lack-thereof. The laboratory experiment described below removes these confounding factors from consideration, thereby allowing us to focus solely on the effects of the four features that concern us.

3.2 Design

To implement such an institution, we build upon the framework and treatments of AH2009 and AH2011 by adding a new treatment. “OPTION2” is a version of AH2009’s “OPTION” treatment wherein both an investor and their counterpart trustee know that the investor has the option to avoid the knowledge of betrayal. In the laboratory, both investors and trustees sit in visually isolated terminals in the same room. All subjects are given the complete instructions of both the investors and the trustees and listen to an experimenter read the instructions out-loud to the entire room (see Appendix for the instructions). This ensures that all subjects are aware of the institutional design that allows investors to be paid based upon the computer draw, thereby shielding them from the knowledge of personal betrayal.[6]

Note that the implemented institution contains each of the four features, from 3.1, that we sought to implement. OPTION2 is designed to ensure that all differences in the results, when compared to the results of AH2009 and AH2011, can be attributed to betrayal aversion. Additionally note that when subjects choose between the “human” and “computer” alternatives in OPTION2, the computer option removes only aversion to knowledge of a *personal* betrayal. Any investor who is averse to the knowledge that *some* trustee in the group betrayed trust is not shielded from that knowledge and thus the subject’s choice reflects the institution’s reduction in the *magnitude* of the emotional disutility associated with betrayal without reducing the *probability* of betrayal.

[6] We highlight the importance of making the environment a Common-Knowledge institution, where all subjects know that all subjects know the full details of the experiment, which is more reflective of naturally existing institutional rules.

3.3 Hypotheses

Hypothesis 1: *There will be a decreasing trend in betrayal from DONTKNOW2 to OPTION2 to KNOW.*

We predict that the willingness to betray will increase as the disutility trustees expect to face from taking a bigger share of the money decreases (due to guilt aversion (Charness and Dufwenberg, 2006), altruism (Fehr, 2009), "moral wiggle-room" (Dana et al. 2007), etc.) In KNOW, an investor who chooses to trust is forced to know whether he/she was personally betrayed. In DONTKNOW2, it is impossible for a trustee's counterpart investor to be exposed to the knowledge of betrayal. Trustees therefore have complete moral wiggle room and there can be no added disutility associated with betraying a betrayal-averse investor. This is the reason, in AH2011, for the increased rates of betrayal in DONTKNOW2 as compared to KNOW. In OPTION2, however, a trustee's decision to betray trust may or may not directly affect their counterpart investor's payoff. The investor chooses whether to be exposed to betrayal knowledge. Therefore, trustees do not have the moral wiggle room present in DONTKNOW2. Similarly, trustees no longer know for sure that their counterpart is shielded from betrayal aversion. Hypothesis 1 arises due to the fact that the cost of betrayal is expected to be greater in OPTION2 than DONTKNOW2, but less in OPTION2 than KNOW.

Hypothesis 2: *Trust in OPTION2 will be significantly higher than trust in DONTKNOW2.*

If trustees display the behavior predicted in Hypothesis 1, then an investor is expected to prefer investing in OPTION2 than DONTKNOW2, due to lower betrayal rates in OPTION2. Also, while betrayal-averse investors can avoid betrayal knowledge in both environments, only in OPTION2 can investors expose themselves to the knowledge of betrayal by trusting if, for example, they get a large positive utility from the knowledge of personal reciprocation of trust. Therefore, the overall expected utility of trusting is expected to be greater in OPTION2 than DONTKNOW2, which is Hypothesis 2.

Hypothesis 3: *Trust in OPTION2 will be significantly greater than trust in KNOW.*

The comparison of KNOW to OPTION2 involves two countervailing factors. The possible increased monetary risk, due to moral wiggle-room and the reduced disutility associated with betraying non-betrayal-averse investors, is expected to lead some investors who would trust in KNOW to not trust in the common-knowledge treatments, thereby reducing observed trust. Nevertheless, the ability, in the common knowledge treatments, to avoid the knowledge of betrayal when trusting, is expected to lead some investors who would not have trusted in KNOW to trust in OPTION2. This will result in an increased frequency of trust. For OPTION2 to be a successful social exchange institution, i.e. for it to reduce the negative impact of betrayal aversion, the increase in trust must be greater than any decrease resulting from any expected increase in betrayal rates. We therefore predict that OPTION2 will result in a higher frequency of trust than KNOW.

4 Results

We report data from 309 students (26 pairs in each of KNOW and OPTION treatments, 28 in DONTKNOW, 32 in OPTION2, and 34 in DONTKNOW2.)[7] Each subject received a seven-dollar show-up bonus in addition to their earnings from the experiment, and spent about 45 minutes in the lab. All sessions occurred at the Interdisciplinary Center for Economic Science (ICES) Laboratory at George Mason University.

4.1 Trustee Behavior

From the trustees' perspective KNOW, KNOW2, OPTION, and DONTKNOW are identical due to the fact that all trustees have the same instructions. Indeed, one cannot dispute that trustees behave statistically similarly in these four treatments (F-test, p=.887), with betrayal rates of 69.2%, 67.8%, 61.5%, and 60% respectively.[8] For analysis of trustee decisions we pool these treatments together and refer to them as the BASELINE treatments with a mean betrayal rate of 65.0%.

[7] We eliminate the data from the 3 investors in DONTKNOW treatments who played in a KNOW or OPTION game (done in AH2009 to ensure no trustees were deceived); we also include the data from the 20 trustees in KNOW2 (the robustness test of KNOW reported in AH2009,) as they made decisions in the exact same environment as the trustees in KNOW, DONTKNOW, and OPTION. The data from these four treatments were previously reported in AH2009, while data from DONTKNOW2 was previously reported in AH2011. The data from OPTION2, the key manipulation in this study, has not been previously reported.

[8] Additionally all pairwise comparisons are insignificantly different (two-tailed Mann-Whitney, p>.500).

Result 1: *There is a significantly decreasing trend in betrayal rates from DONTKNOW2 to OPTION2 to the BASELINE treatments.*

Consistent with Hypothesis 1, and as illustrated by Figure 2, as the cost of betrayal increases the rate of betrayal decreases (p = 0.034, Cuzick trend test), from 85.3% in DONTKNOW2 to 68.8% in OPTION2 to 65.0% in BASELINE.

<Figure2>

4.2 Investor Behavior

Result 2: *Trust in OPTION2 is significantly higher than trust in DONTKNOW2.*

As predicted by Hypotheses 2, in OPTION2, where investors have the option to avoid the knowledge of betrayal, the frequency (87.5%) of investors trusting is significantly greater than when the institution forces everyone to avoid the knowledge of betrayal in DONTKNOW2, (58.8%, p <0.01).

Result 3: *Trust in OPTION2 is significantly higher than trust in KNOW***.**

As predicted by Hypotheses 3, the 87.5% trusting in OPTION2, is also significantly greater than when there is no institution removing betrayal knowledge (KNOW, 65.0%, p<0.05) Figure 3 illustrates the levels of trust in the KNOW, OPTION2, and DONTKNOW2 treatments from results 2 and 3.

<Figure 3>

Result 4: *The fraction of investors who trust and chose the computer-trust option, in OPTION2 is not significantly greater than in OPTION.*

Figure 4 shows that the fraction of subjects in the OPTION2 treatment trusting the computer increased from 53.9% to 59.4%, but the increase is not significant (p > 0.6). The marginal increase in the fraction not trusting (from 0% in OPTION to 12.5% in OPTION2, p=0.064) can be attributed to a decrease in the fraction trusting their counterpart when the institution becomes common knowledge (decreasing from 46.2% to 28.1%); however, this change is not significant (p = 0.159).

<Figure 4>

These results suggest that investors account for the changes in trustee behavior observed in Result 1. The significant increase in trust in OPTION2, as compared to KNOW, is consistent with both the reduction of the emotional/psychological risk of betrayal and a belief that trustee betrayal would not increase significantly when trustees know that investors are able (but not required) to avoid betrayal knowledge.

5 Conclusions

Institutions implemented by firms, corporations, governments, or even small groups of people, are complex combinations of rules. We tested the simple rule that an investor has the option, but not the requirement, to avoid knowledge of personal betrayal. We found that an institution providing such a rule not only maintains the lower rates of betrayal seen when betrayal aversion influences agent behavior, but also results in significantly increased trust rates.

Institutions can affect betrayal aversion in multiple ways, e.g., by affecting the probability that a trusting agent will experience betrayal or by reducing the magnitude of the disutility associated with experiencing betrayal. To our knowledge, the present study is the first to examine the latter. The importance of distinguishing between these two effects is highlighted by past studies suggesting that reducing the probability of experiencing betrayal could increase disutility if a betrayal occurs. Intuitively, the unexpected betrayal of a close friend is a more emotional event than the betrayal of a stranger. This effect also translates into consumer relations. For example, Grégoire and Fisher (2008) demonstrate that when loyal customers, who expect less betrayal, are betrayed, they retaliate more aggressively than less loyal customers. Thus, institutions that decrease social distance or reduce anonymity, such as reputation effects, repeat interactions, or communication, may actually serve to increase the magnitude of emotional disutility associated with experiencing betrayal. Such examples illustrate the need for institutions to include some means of reducing the disutility when betrayal occurs.

Our research emphasizes the importance of "moral wiggle room" (Dana et al., 2006; Dana et al., 2007) in social exchange influenced by betrayal aversion (AH 2011). These past studies suggest that increasing moral wiggle room in a trust environments is detrimental to exchange in that it can lead to significant increases in betrayal. Making betrayal knowledge optional in OPTION2 creates an unique moral environment since whether there is moral wiggle room is probabilistic. Our data provides evidence on how trustees act with this type of probabilistic moral wiggle room. Additionally, since differences between our treatments reflect differences in the presence of betrayal knowledge (e.g. moral wiggle room), the difference between expected earnings from the BASELINE treatments (E(Earnings)= \$6.55) in each treatment can be thought to reflect the value of "moral wiggle room" to trustees. When the institution removes all betrayal aversion from the environment, in DONTKNOW2, the value of moral wiggle room is shown to be significant, \$2.68 (two-tailed Mann-Whitney, p<0.01) as the expected earnings fall from \$6.55 to \$4.00. However, when the institution makes the removal of betrayal knowledge only optional, the probablisitic increase in moral wiggle room is seen to be worth little to trustees. This is reflected by the insignificant increase in betrayal rates from BASELINE to OPTION2 where expected earnings drop an insignificant 49 cents (p>0.69) from BASELINE to \$6.06). The lack of a significant increase in betrayal rates suggests that institutions with optional removal of betrayal knowledge avoid the adverse effects of full moral wiggle room (significantly increased betrayal) as found in DONTKNOW2.

Additionally, we contribute to recent work that demonstrates how natural factors contribute to the efficiency of exchange environments. Naturally-evolved factors have been shown to provide the "solution" to many laboratory environments where subject behavior appears inefficient. For example, Xiao and Houser (2005) show that the reintroduction of communication (emotion expression) substantially reduces the rejection of unfair offers in ultimatum games. Fehr and Gachter (2000) and Masclet et al. (2003) show that monetary and informal punishment (normally available in some form in natural environments) can substantially increase cooperation in public goods environments. Likewise, guilt aversion and promises (Charness and Dufwenberg, 2006) improve reciprocation in trust games. These studies provide evidence that inefficiencies sometimes observed in the laboratory may be partially attributable to the inability of participants to employ naturally-evolved factors.

In undesigned (so called, "real world") environments, the move from personal to impersonal exchange creates a barrier for natural factors, including betrayal aversion. For example, the switch from "mom and pop" stores to chain stores removed the knowledge of betrayal (by a specific person) from economic exchange (e.g., one does not usually feel betrayed by a checkout clerk after purchasing a product with unexpectedly low quality.) Many companies appear to compensate for this by creating new channels for betrayal knowledge to enter the exchange environment. For example, with customer service phone lines, people can choose to inquire about an adverse purchase outcome and thus discover if the betrayal was personal. Note that this institution provides an "option" to discover whether one was betrayed, in the same spirit as we tested above.

Closer yet to our study, many shipping companies provide detailed tracking information about the status of a shipment. Clearly, a customer has the option to observe or ignore this information. The shipping company's employees generally do not know which customers will obtain this information, but they do know this information exists. Moreover, they do not know how betrayal aversion influences the decisions of their customers. While these rules serve many functions, our study illustrates that they can be expected to have the (possibly unintended) consequence of increasing overall trust (willingness to ship), and thus may partially explain why this institution has survived.

Studies like ours and Aimone and Houser (2011) suggest that customers may receive low-quality products from companies that shield them from betrayal knowledge. Perhaps a combination of return policies and quality guarantees could effectively compensate for the removal of betrayal aversion. Future studies in this direction would be valuable, particularly if they account for the impact of such rules and policies on betrayal aversion.

References

Aimone, J.A., and D. Houser (2011) Beneficial Betrayal Aversion. *PLoS ONE* 6(3): e17725. doi:10.1371/journal.pone.0017725

Aimone, J. A., and D. Houser (2009): What You Don't Know Won't Hurt You: A Laboratory Analysis of Betrayal Aversion, Working Paper.

Berg, J., J. Dickhaut, K. McCabe (1995): Trust, Reciprocity, and Social History. *Games and Economic Behavior*, 10, 122-142.

Birnberg, J. G. and Y. Zhang (2010) When Betrayal Aversion Meets Loss Aversion: The Effect of Economic Downturn on Internal Control System Choices. *Journal of Management Accounting Research*, Forthcoming.

Bohnet, I., F. Grieg, B. Herrmann, and R. Zeckhauser (2008): Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *The American Economic Review* 98(1), 294-310.

Bohnet, I. and R. Zeckhauser (2004): Trust, Risk and Betrayal. *Journal of Economic Behavior and Organization*, 55, 467-484.

Bolton, G. E. and A. Ockenfels (2010): Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States: Comment. *American Economic Review, 100(1), 628–633.*

Charness, G., and M. Dufwenberg (2006): Promises and Partnership. *Econometrica,* 74(6), 1579-1601.

Dana, J., R.A. Weber, J.X. Kuang (2007): Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness. *Economic Theory*, 33: 67–80.

Dana, J., D.M. Cain, R. M. Dawes (2006): What You Don't Know Won't Hurt Me: Costly (but Quiet) Exit in Dictator Games. *Organizational Behavior and Human Decision Processes*, 100, 193 – 201.

Falk, A., E. Fehr, and U. Fischbacher (2008): Testing Theories of Fairness – Intentions Matter. *Games and Economic Behavior,* 62: 287-303.

Fehr, E. (2009): On the Eonomics and Biology of Trust. *Journal of the European Economics Association*, 7(2-3)235-266.

Fehr, E. and K.M. Schmidt (1999): A Theory Of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics,* 114(3), 817-868.

Fehr, E. and S. Gachter (2000): Cooperation and Punishment in Pub- lic Goods Experiments. *American Economic Review*, 90(4), pp. 980-94.

Fehr, E., and U. Fischbacher (2003): The Nature of Human Altruism. *Nature* 425: 785-791

Gross, J. (1998): The Emerging Field of Emotion Regulation: An Integrative Review. *Review of General Psychology*, 2(3) 271-299.

Hong, K. and I. Bohnet (2007): Status and Distrust: The Relevance of Inequality and Betrayal Aversion. *Journal of Economic Psychology*, 28, 197-213.

Houser, D., D. Schunk, and J. Winter (2008): Trust Games Measure Trust. Working paper.

Houser, D., and J. Wooders (2006): Reputation in Auctions: Theory, and Evidence from eBay, *Journal of Economics & Management Strategy*, 15(2), Summer 2006, 353–369.

Koehler, J. J., and A. D. Gershoff (2003): Betrayal Aversion: When Agents of Protection Become Agents Of Harm. *Organizational Behavior and Human Decision Processes*, 90, 244-261.

Koehler, J. J., and A. D. Gershoff (2011): Safety First? The Role of Emotion in Safety Product Betrayal Aversion. *Journal of Consumer Research*, (Forthcoming)

Kosfeld, M., M. Heinrichs, P. J. Zak, U. Fischbacher, and E. Fehr (2005): Oxytocin Increases Trust in Humans. *Nature*, 435, 673-676.

Masclet, D., C. Noussair, S. Tucker, M. Villeval (2003): Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism, *The American Economic Review*, 93(1) , 366-380.

Miu, A. C., R. M. Heilman, D. Houser (2008): Anxiety Impairs Decision-making: Psychophysiological Evidence from an Iowa Gambling Task. *Biological Psychology*, 77, 353-358.

Ochsner, K.N., J.J. Gross (2005): The Cognitive Control of Emotion. *TRENDS in Cognitive Sciences*, 9(5) 242-249.

Tricomi, E., A. Rangel, C.F. Camerer, J.P. O’Doherty (2010): Neural Evidence for Inequality-Averse Social Preferences. *Nature* 463: 1089-1091.

Xiao, E and D. Houser (2005): Emotion Expression in Human Punishment Behavior, *Proceedings of the National Academy of Sciences of the United States of America,* 102(20), 7398-7401.

Yamagihi, T., Y. Horita, H. Takagishi, M. Shinada, S. Tanida, and K.S. Cook (2009): “The Private Rejection of Unfair Offers and Emotional Commitment”, *Proceedings of the National Academy of Sciences of the United States of America,* 106(28), 11520-11523.

Figure 1
Investment Game

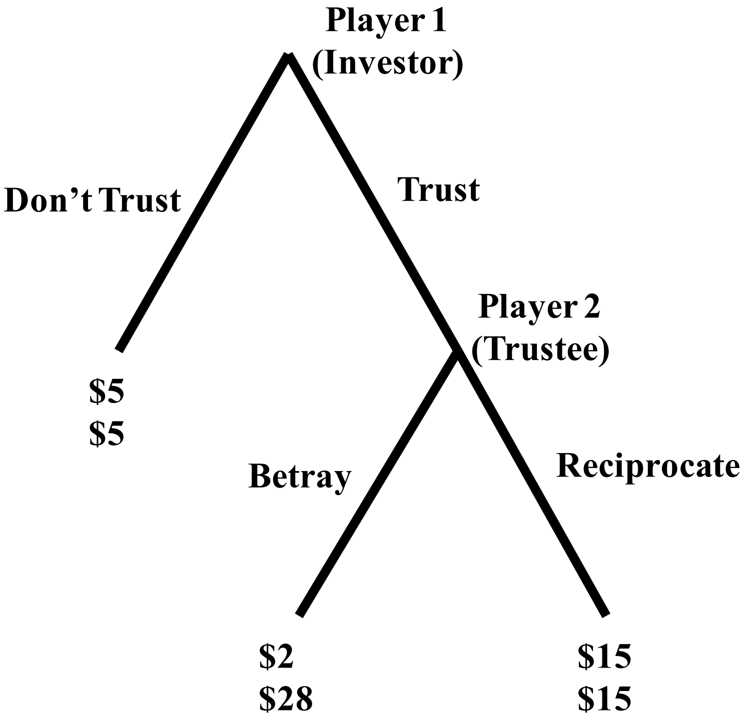


Figure 2
Percent Choosing to Betray
By Treatment

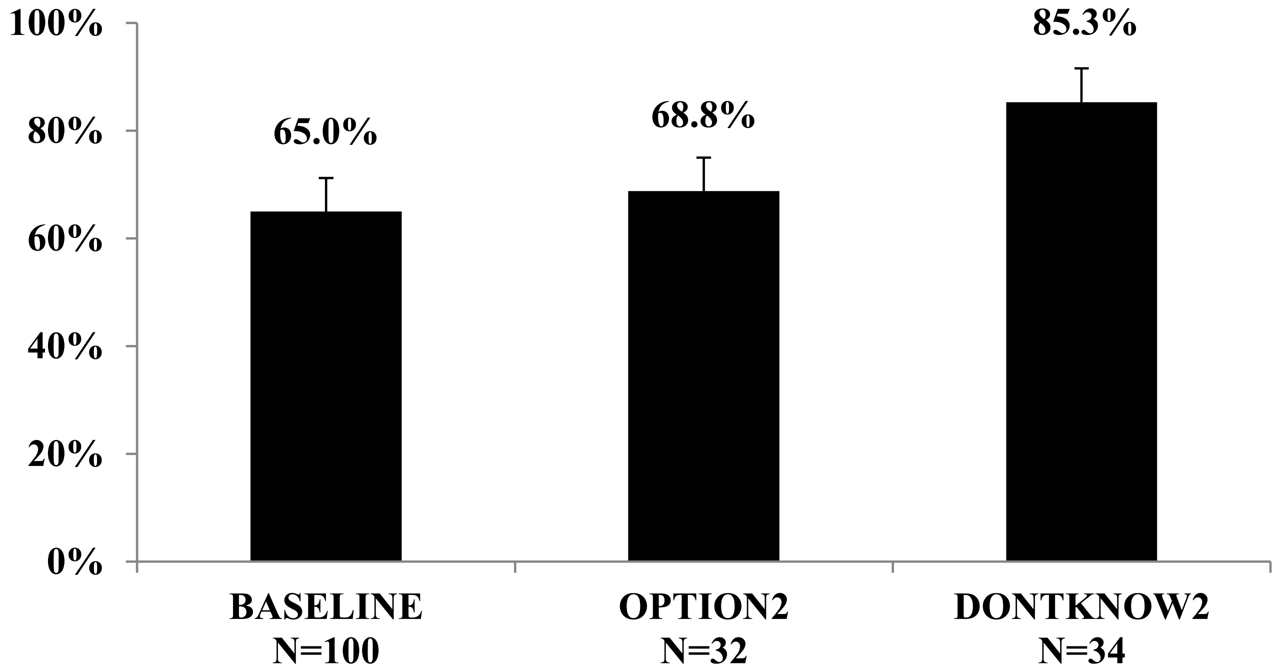


Figure 3
Percent Choosing to Trust
By Treatment

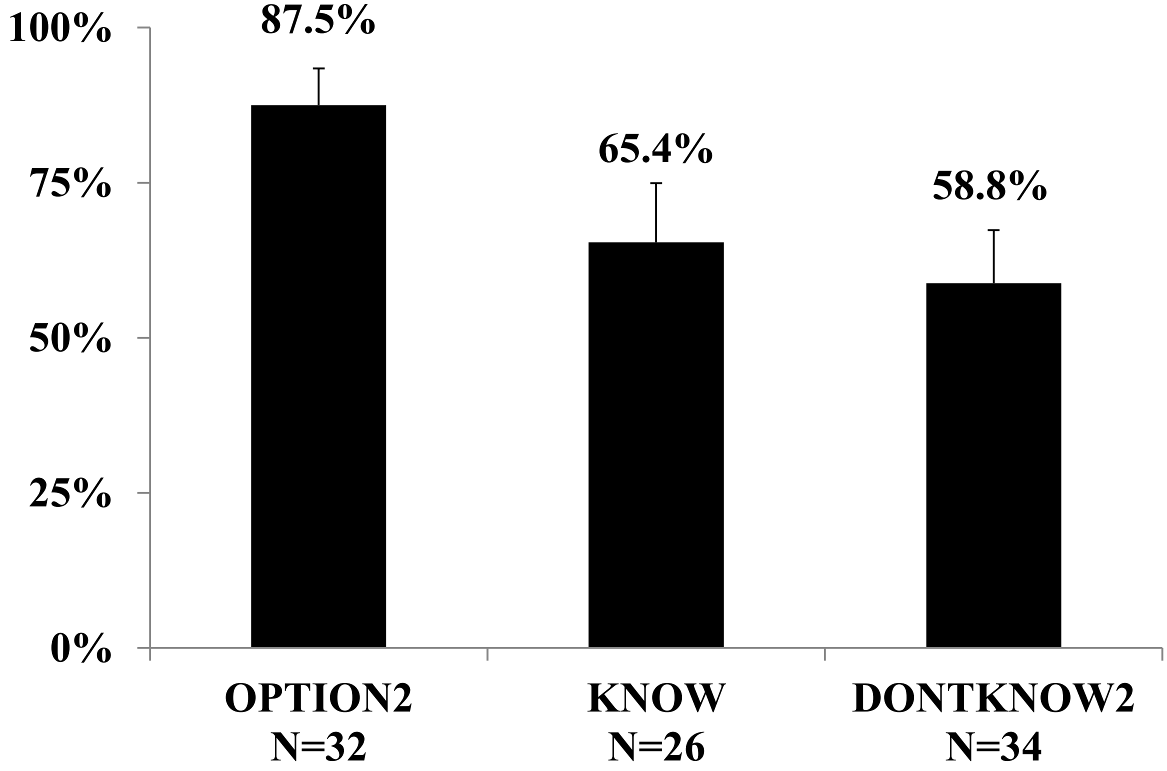
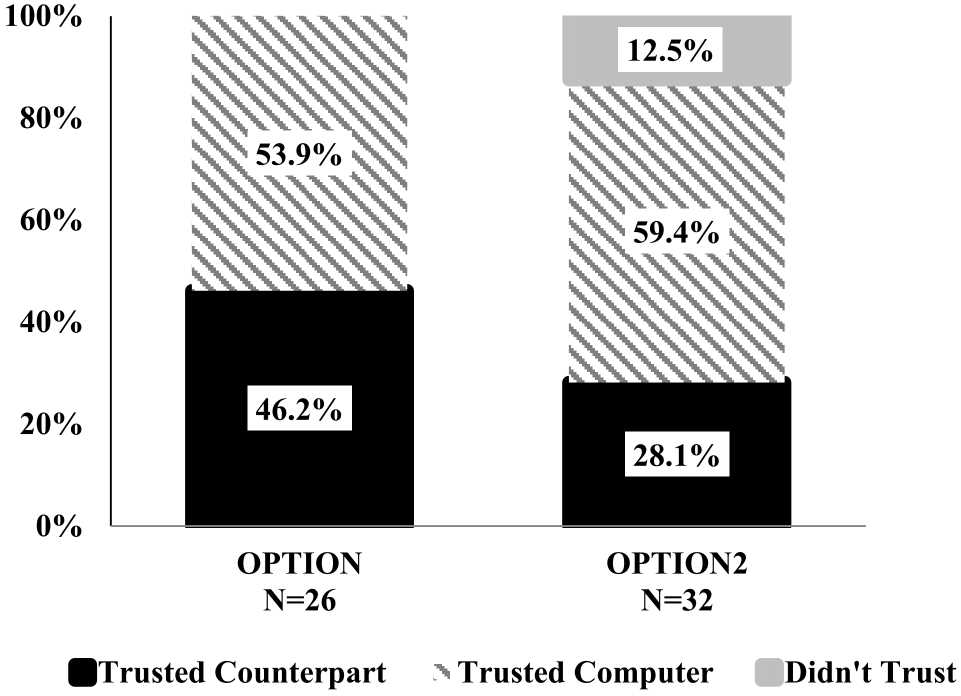


Figure 4
OPTION Treatment Choices



Appendix A
Room A (Investors) Instructions: OPTION Treatment[9]

Thank you for participating in today's experiment. You've earned a $7 show-up bonus for participating. In reading and following the instructions below, you have the potential to earn significantly more. You have been randomly assigned to **Room A.** You will also be randomly and anonymously assigned to a person in **Room B.** Your counterpart will not be told your name, and you will not be told his/her name.

How you are matched with your counterpart:
Each of the 10 Room A persons will be matched with a different Room B counterpart for the entire experiment. The experimenter will bring around a box with the numbers 1 through 10 inside. The number you draw will assign you to one of the 10 counterparts in Room B (B1 through B10 coinciding with the numbers 1 through 10 in the box). The number also matches you with one of the 10 computer number decisions (coinciding with numbers 1 through 10 in the box).

Your Decision:
You have three options for how the earnings for you and your counterpart will be determined in today's experiment. You must choose exactly one of the following three options:

- You receive $5 and your counterpart receives $5.
- Both you and your counterpart are paid based on his/her decision between **"U"** ($15 for you and $15 for him/her) and **"D"** ($2 for you and $28 for him/her).
- Your counterpart is paid according to his/her decision between **"U"** and **"D"**, and you are paid based on a computer's choice between either **"U"** or **"D"**.

You will not be told what the computer's decision was, or what your counterpart's decision was, unless you choose that earnings option.

[9] DONTKNOW treatment did not include the second payment choice option, KNOW treatment did not include third payment choice option and did not include the "computer's decision" paragraph (next page.) OPTION2 instructions are identical to the OPTION instructions from Appendix A. The only way that OPTION2 differs is that both subjects in the role of Investors and subjects in the role of Trustees received both room A and room B instructions.

Room B Decision: (The instructions given to your counterpart)
You will be anonymously assigned to a Room A counterpart who drew your number randomly from a box with the numbers 1 through 10 inside. This person will be your counterpart for the entire experiment. Your counterpart will make a decision that can affect your earnings in today's experiment. He or she can choose for both of you to be paid $5. Another possibility is that he/she will let you determine both of your payoffs. If he/she chooses this option and you choose **"U"**, then you get paid $15 and he/she gets paid $15. If you choose **"D"**, then you get paid $28 and he/she gets paid $2. Your payoff will be determined in one of these two ways. Your counterpart can choose only one of the earnings methods. We will ask you to make your decision on **"U"** or **"D"** at the same time that your counterpart makes his or her choice. Your decision will only determine your payoff if your counterpart did not choose the option to give you $5.

Computer's Decision:
After the Room B participants make their decisions, the computer will assign either **"U"** or **"D"** to each of the ten numbers. The computer has been programmed to assign dollar values to each of the 10 numbers in the box according to the decisions made by the Room B participants. What this means is that the number of **"U"** choices made by the computer is exactly the same as the number of **"U"** choices made by the participants in room B. Also, the number of **"D"** choices made by the computer is exactly the same as the number of **"D"** choices made by the room B participants. (Note: while the number of **"U"** numbers and number of **"D"** numbers are the same as in the Room B decisions, which numbers are assigned **"U"** or **"D"** is randomly decided by the computer) For example: if five Room B participants choose **"U"**, then five of the numbers between 1 and 10 are randomly assigned to have the **"U"** payoff, and the remaining five numbers are assigned to the **"D"** payoff. (Note: the numbers used here are only an example and not necessarily representative of Room B decisions)

Room B (Trustees) Instructions

Thank you for participating in today's experiment. You've earned a $7 show-up bonus for participating. In reading and following the instructions below, you have the potential to earn significantly more. You have been randomly assigned to **Room B.** You will also be randomly and anonymously assigned to a person in **Room A.** Your counterpart will not be told your name, and you will not be told his/her name.

You will be anonymously assigned to a Room A counterpart who drew your number randomly from a box with the numbers 1 through 10 inside. This person will be your counterpart for the entire experiment. Your counterpart will make a decision that can affect your earnings in today's experiment. He or she can choose for both of you to be paid $5. Another possibility is that he/she will let you determine both of your payoffs. If he/she chooses this option and you choose **"U"**, then you get paid $15 and he/she gets paid $15. If you choose **"D"**, then you get paid $28 and he/she gets paid $2. Your payoff will be determined in one of these two ways. Your counterpart can choose only one of the earnings methods. We will ask you to make your decision on **"U"** or **"D"** at the same time that your counterpart makes his or her choice. Your decision will only determine your payoff if your counterpart did not choose the option to give you $5.

Appendix B
Instructions (All Subjects): DONTKNOW2 Treatment[10]
Room A Instructions

Thank you for participating in today's experiment. You've earned a $7 show-up bonus for participating. In reading and following the instructions below, you have the potential to earn significantly more. You have been randomly assigned to **Room A.** You will also be randomly and anonymously assigned to a person in **Room B.** Your counterpart will not be told your name, and you will not be told his/her name.

How you are matched with your counterpart:
Each of the 10 Room A persons will be matched with a different Room B counterpart for the entire experiment. The experimenter will bring around a box with the numbers 1 through 10 inside. The number you draw will assign you to one of the 10 counterparts in Room B (B1 through B10 coinciding with the numbers 1 through 10 in the box). The number also matches you with one of the 10 computer number decisions (coinciding with numbers 1 through 10 in the box).

Your Decision:
You have two options for how the earnings for you and your counterpart will be determined in today's experiment. You must choose exactly one of the following two options:

- You receive $5 and your counterpart receives $5.
- Your counterpart is paid according to his/her decision between **"U"** and **"D"**, and you are paid based on a computer's choice between either **"U"** or **"D"**

You will not be told what the computer's decision was unless you choose that earnings option.

[10] OPTION2 instructions are identical to the OPTION instructions from Appendix A. The only way that OPTION2 differs is that both subjects in the role of Investors and subjects in the role of Trustees received both room A and room B instructions.

Room B Decision: (The instructions given to your counterpart)
You will be anonymously assigned to a Room A counterpart who drew your number randomly from a box with the numbers 1 through 10 inside. This person will be your counterpart for the entire experiment. Your counterpart will make a decision that can affect your earnings in today's experiment. He or she can choose for both of you to be paid $5. Another possibility is that he/she will let you determine both of your payoffs. If he/she chooses this option and you choose **"U"**, then you (room B person) get paid $15. If you choose **"D"**, then you (room B person) get paid $28. Your payoff will be determined in one of these two ways. Your counterpart can choose only one of the earnings methods. We will ask you to make your decision on **"U"** or **"D"** at the same time that your counterpart makes his or her choice. Your decision will only determine your payoff if your counterpart did not choose the option to give you $5.

Computer's Decision:
After the Room B participants make their decisions, the computer will assign to each of the ten numbers either **"U"**, meaning a payoff of $15 to you (room A person), or **"D"**, meaning a payoff of $2 to you (room A person). The computer has been programmed to assign dollar values to each of the 10 numbers in the box according to the decisions made by the Room B participants. What this means is that the number of **"U"** choices made by the computer is exactly the same as the number of **"U"** choices made by the participants in room B. Also, the number of **"D"** choices made by the computer is exactly the same as the number of **"D"** choices made by the room B participants. (Note: while the number of **"U"** numbers and number of **"D"** numbers are the same as in the Room B decisions, which numbers are assigned **"U"** or **"D"** is randomly decided by the computer) For example: if five Room B participants choose **"U"**, then five of the numbers between 1 and 10 are randomly assigned to have the **"U"** payoff of $15, and the remaining five numbers are assigned to the **"D"** payoff of $2. (Note: the numbers used here are only an example and not necessarily representative of Room B decisions)

Room B Instructions

Thank you for participating in today's experiment. You've earned a $7 show-up bonus for participating. In reading and following the instructions below, you have the potential to earn significantly more. You have been randomly assigned to **Room B.** You will also be randomly and anonymously assigned to a person in **Room A.** Your counterpart will not be told your name, and you will not be told his/her name.

You will be anonymously assigned to a Room A counterpart who drew your number randomly from a box with the numbers 1 through 10 inside. This person will be your counterpart for the entire experiment. Your counterpart will make a decision that can affect your earnings in today's experiment. He or she can choose for both of you to be paid $5. Another possibility is that he/she will let you determine both of your payoffs. If he/she chooses this option and you choose **"U"**, then you get paid $15. If you choose **"D"**, then you get paid $28. Your payoff will be determined in one of these two ways. Your counterpart can choose only one of the earnings methods. We will ask you to make your decision on **"U"** or **"D"** at the same time that your counterpart makes his or her choice. Your decision will only determine your payoff if your counterpart did not choose the option to give you $5.