Global Research Unit Working Paper #2020-019

Design Your Trustworthiness

Sen Geng, Xiamen University
Menglong Guan, UC Santa Barbara

Design Your Trustworthiness*

Sen Geng
Xiamen University

Menglong Guan
UC Santa Barbara

Abstract

One realistic market innovation to tackle social dilemma involving trust is that trustees take the lead in fostering trust by designing information about trustworthiness. We propose two novel experimental games to study the causal effect of such information design on trustworthiness and trust, and also to explore the underlying mechanism. Experimental data from a within-subject design shows that the treatment effect is generally consistent with equilibrium analysis in terms of direction but not magnitude, and several behavioral patterns deviate considerably from the prediction. We then propose a model allowing for subjects' heterogeneity in prosociality and strategic sophistication that rationalizes the data. We apply the maximum likelihood estimation method to estimate each subject's behavioral type and find that the estimated type almost fully coincides with the type prescribed by the model and is largely consistent with the type assigned by an intuitive criterion. We finally provide evidence that prosociality and strategic sophistication is orthogonal.

Keywords: trust, trustworthiness, information design, prosociality, strategic sophistication

---

*Sen Geng: sg1670@xmu.edu.cn; Menglong Guan: mguan@ucsb.edu.

JEL: C92, D82, D83, D91, M31

Highlights:

1. We explore a realistic market innovation to tackle social dilemma involving trust.

2. An introduction of information design about trustworthiness boosts trustworthiness and trust.

3. 28% (26%) of the time the least (most) Blackwell informative structure is employed.

4. Prosociality and strategic sophistication are orthogonal.

5. A model of heterogeneity in prosociality and strategic sophistication rationalizes the data.

1 Introduction

Trust is a crucial ingredient for most of social interactions and economic transactions.[1] Trusting act is socially desirable and beneficial to trustees.[2] Trusting act also benefits trustors if trustees are truthworthy but otherwise it is likely to make trustors worse off. Foreseeing the risk of being exploited by trustees, trustors are discouraged to place trusting act in many settings. Clearly, if trust can be built through some innovation then seemingly impossible transactions become possible and all parties involved become better off.

In practice, innovations from different aspects are came up with and carried out to tackle the social dilemma involving trust. The applicability and the outcome of these innovations is largely context specific. For instance, a first solution to the dilemma is simply make a formal contract between two parties. This approach resolves the dilemma only when relevant interests can be completely contracted and the contract can be enforced in an economical way.[3] A second solution is the reputation incentive. Trustees' consideration of long-run interests from others' trusting act is likely to prevent them from exploiting trustors' trust. The working of such reputation mechanism typically relies on repeated interactions, verifiable reputation and trustees' believing in the monotonicity between reputation and long-run interests. A third solution is a market solution involving a third party like entrepreneurs or credit reporting agencies who operate a business from facilitating

[1] For the positive impact of trust on economic outcomes, see Guiso et al. (2004, 2006, 2008, 2009), Knack and Keefer (1997), and Zak and Knack (2001), among others.

[2] The Oxford English Dictionary defines trust to be the belief that sb/sth is good, sincere, honest, etc. and will not try to harm or trick you. Nevertheless, existing economic studies of trust typically use a behavioral definition that is based on Coleman (1990) for the reasons that measures of act have been better developed than measures of belief and that the behavioral definition helps researchers to understand trusting act in terms of preferences and beliefs (Fehr, 2009). The link between the belief definition and the behavioral definition is investigated in Glaeser et al. (2000). Our article implicitly adopts the behavioral definition. Notably, since most of our analyses are at the individual level and individual preference is arguably stable, our conclusions still remain under the belief definition as long as we interpret observed act as a proxy for belief.

[3] An enforcement is economical for one party if the cost of trustees' breaking the contract outweighs its benefit or the benefit of trustors' seeking contract enforcement covers its cost.

trusting act. A well-known example of such solution is online marketplaces such as eBay and Taobao where the money paid by a buyer is withheld by the platform and is not transferred to a seller until paid goods and services are shipped to or received by the buyer.[4] A fourth solution is a market solution that trustees take the lead in fostering trust and make every effort to gain trust from trustors.[5] In particular in business practices involving seller and buyer or investee and investor, trustees are highly motivated to take initiatives to build trust.

Of those initiatives trustees take to gain trust, information design is a rather common approach, especially in the realm of equity investments, and financial products and services. For instance, a financial product is usually designed and marketed in either of two opposite ways. In one extreme, it is designed in a plain and much informative way so that even layman investors can easily appreciate the essence and in turn trust the product. In the other extreme, (perhaps intentionally) it is designed intricately and uninformatively so that even savvy investors have trouble in figuring it out but nevertheless place trusting act.[6] In this article we ask the following questions: Does the common practice of designing information to gain trust indeed boost trusting act? What's the mechanism underlying the effects/null effects of such information design?

To address these questions, we introduce experimental games that abstract features about trust, trustworthiness and information design from the multifaceted features of practices.[7] In our motivating example about financial products, it is a fact that both the asocial risk per se of a financial product and the social risk about investees' trustworthiness are likely to matter for investors' decision. Besides,

[4]This greatly resolves the trust dilemma between remote buyers and sellers and makes many previously impossible transactions become possible.

[5]Efforts trustees initiate to gain trust include: free trial, refund guarantee, product warranties and returns, advertisement and being sponsored by celebrities, and information design about trustworthiness discussed in this article, etc.

[6]A well-known exmaple in this regard is Bernard Madoff's crafting a massive Ponzi scheme for decades.

[7]Karlan (2005) provides evidence that experimental trust games are directly linked to real-life decisions including financial decisions.

these two factors are often of importance to a similar degree but sometimes trust is pivotal to investors' decision. Given the focus of this article, our experimental games do not intend to fully mimic these practices but are designed to highlight the interaction between trust, trustworthiness and information design in these practices while leaving out other features such as the asocial risk.

The two games we introduce are inverse trust game (Game 1) and trustworthiness design game (Game 2).[8] In both games each of two players is endowed with 10 tokens and a trustee moves first to decide which of two allocation plans is used when an investment project is carried out. Both trustee and trustor receive 15 tokens according to the allocation plan $(15, 15)$, and they receive 22 tokens and 8 tokens respectively according to the allocation plan $(22, 8)$. A trustee decides the allocation plan in an indirect way, that is, he/she chooses the chance $p \in [0, 1]$ that plan $(15, 15)$ is to be used. A trustor moves second to decide whether the investment project is carried out. In the case of carrying out the project both players' endowments are spent and they receive payoff according to the allocation plan that is used, and in the case of not carrying out the project both players retain their endowments. In Game 1, a trustor knows neither the trustee's choice of p nor the allocation plan that is used when making the decision. In Game 2, when deciding p a trustee also chooses the conditional likelihoods of generating a signal indicating the allocation plan. More precisely, a signal of either black ball or while ball is generated with probability q_1 for being a black ball when the allocation plan is $(15, 15)$ and with probability q_2 when the allocation plan is $(22, 8)$; and the trustee decides both q_1 and q_2. A trustor in Game 2 observes q_1, q_2 and a signal while still not knowing either the trustee's choice of p or the allocation plan that is used when making his/her decision.

We adopt the notion of Blackwell's informativeness to order information structures (q_1, q_2) in Game 2 (Blackwell, 1951, 1953). We provide a characterization

[8] See in Section 3.1 the discussion about detailed game features.

of Blackwell's order including an equivalent condition about nested likelihood ratios that are easily applied to data. While there is no numerical representation of Blackwell's order in the standard sense, we derive an index that captures a property about Blackwell's order to differentiate information structures. Subgame perfect equilibrium analysis of Game 2 under the standard assumptions reveals that only certain conditional trust equilibria exists: a trustee chooses an information structure such that the likelihood ratio $\frac{q_1}{q_2}$ or $\frac{1-q_1}{1-q_2}$ exceeds a cutoff value and chooses to be trustworthy with positive chance up to 100% chance, and a trustor invests in the project only when both information structure and signal generated from the information structure favors the allocation plan $(15, 15)$. In particular, a trustee's choosing the most Blackwell informative structure and choosing to be fully trustworthy is part of an equilibrium. Compared to the unique equilibrium in Game 1 that a trustee chooses not to be trustworthy and a trustor does not invest in the project, the equilibrium analysis predicts that trustworthiness and trust increase from Game 1 to Game 2.

We design a theory-guided laboratory experiment and employ a within-subject design in which subjects play both Game 1 and Game 2. Experimental results show that the treatment effects are generally consistent with equilibrium analysis in terms of direction but not magnitude, and several behavioral patterns deviate considerably from equilibrium predictions. First, the mean value of trustees' chosen chance of plan (15, 15) being used increases by 87.5% from 0.232 to 0.435 and the fraction of time trustors invest in the project increases by 39.9% from 0.336 to 0.470, both of which are confirmed by various statistical tests. Second, trustees' actions in Game 2 are consistent with equilibria prescriptions 24.25% of the time and interestingly these consistent actions all belong to the equilibrium with full trustworthiness and the most Blackwell informative structure. Third, trustors' actions on the above equilibrium path are largely consistent with the equilibrium predicted action and their actions on the off-equilibrium path are inconsistent with the prediction for a considerable fraction of time. Fourth, at the aggregate level trustees choose the most Blackwell informative structure 26% of the time and surprisingly choose the least one 28% of the time; at the individual level 97.5% of trustees' mean values of p in Game 2 increase from or at least be no less than in Game 1 and their choice of information structure is tightly associated with their choice of trustworthiness. Fifth, it seems that trustors are able to make inference about the underlying state based on signal but a considerable fraction of them fail to realize that information structure per se also conveys a message about trustworthiness.

To account for the documented treatment effects and behavioral patterns we then propose a behavioral model that relaxes the standard assumptions about selfishness and equilibrium model. The behavioral model assumes that trustees and trustors belong to either a standard selfish type or a prosocial type who receives an additional psychic utility from choosing the allocation plan $(15, 15)$ that benefits trustors and from placing trusting act on trustees respectively. Since our experiment focuses on subjects' initial responses with learning being greatly suppressed, we also assumes that both players are heterogeneous in their levels of strategic sophistication following a structure of the level-k model.[9] Additionally, we assume that trustors hold a heterogeneous viewpoint about the fraction of prosocial trustees in the population, which together with trustors' different levels of reasoning determine their prior about the allocation plan $(15, 15)$ before observing an information set. We finally assume that trustors hold a conditionally pessimistic posterior and in turn employ a conditional maxmin strategy when observing zero-probability information sets, i.e., when Bayes' law is simply not defined.[10] The

[9]See Nagel (1995), Stahl and Wilson (1994, 1995), and Camerer et al. (2004) for early contributions of the level-k model. See Crawford and Iriberri (2007a,b) for its applications to games with incomplete information. See Crawford et al. (2013) for a recent survey on its empirical and theoretical literature.

[10]See Ortoleva (2012) for an alternative model that characterizes non-Bayesian reactions to zero-probability event or small-probability event.

behavioral model equipped with these assumptions enables to generate predictions that are in line with all observed patterns. In particular, it rationalizes some trustees' choosing the least Blackwell informative structure with a zero level of trustworthiness.

Applying the structure imposed by the behavioral model, we employ the maximum likelihood estimation method to identify each subject's behavioral type. Further likelihood ratio tests at the individual level reject the null hypothesis of a random choice model at prespecified significance levels for the majority of subjects, e.g., 89.4% of trustees and 79.8% of trustors at the five percent significance level. We also compare each subject's type based on econometric estimation with his/her type based on an exact criterion from the model (if applicable) and find the two types almost always coincide. In addition, we propose an intuitive criterion, which lifts the very stringent requirement of the exact criterion but maintains some of its intuitive features, to classify a subject's type. We find that the econometric type is largely in line with the type based on the intuitive criterion, that is, for over 75% of trustees and over 92% of trustors (if applicable). The high consistency shows that econometric estimation indeed confirms intuitive analysis of subjects' types and additionally refines our analysis when intuition is not able to provide a clear guidance.

Once a subject is assigned a behavioral type based on econometric estimation, we know his/her characteristics in the following dimensions: prosociality, strategic sophistication and viewpoint about trustees' prosociality if the subject is a trustor. We then explore statistical independence between the two or three dimensions. We find that a trustee's prosociality is not associated with his/her level of strategic sophistication and neither is a trustor. We also find that a trustor's viewpoint about trustees' prosociality is positively associated with his/her level of strategical sophistication while it is not associated with the trustor's prosociality. The findings indicate that prosociality is orthogonal to strategic sophistication while the first-order belief about prosociality and strategic sophistication may coevolve.

Outline. This article is organized as follows. Section 2 discusses the related literature to which our research contributes. Section 3 introduces experimental games and explores informativeness ordering and equilibrium analysis in these games. Section 4 presents experimental design and procedure, followed by experimental results in Section 5. Section 6 proposes a behavioral model, based on which type estimation and orthogonality analysis are provided in Section 7. Section 8 concludes. Appendix A, B, and C provide all proofs and additional propositions, additional data analysis, and experimental instructions respectively.

2 Contributions to the literature

Our article contributes to the literature concerning trust. Existing studies of trust games (mostly being variants of Berg et al. (1995)) share a similar game paradigm in the sense that trustors take the lead by moving first and trustees play a follower role by responding to trustors' actions. Our proposed trust games set up a new paradigm by asking trustees to take the lead in their strategic interaction with trustors. This paradigm change allows us to capture some realistic practices about trust in which trustees play a leading role, and more importantly it opens a new door for understanding what drives trusting act in addition to those already explored in the paradigm of trustors' moving first. With this innovative perspective, we model a realistic and voluntary market innovation to boost trusting act as a combination of inverse trust game and information design game. This is novel to the literature that explores how to resolve the social dilemma involving trust. The existing literature in this regard mainly proposes mechanisms that are related to reputation incentive, cheap-talk communication, and competition in the paradigm of trustors' moving first (Bracht and Feltovich, 2009; Brown et al., 2004; Charness and Dufwenberg, 2006; Huck et al., 2012).[11] Besides, Andreoni (2018) argues that satisfaction guarantee is a market innovation to promote trust and reduce moral hazard, and models the practice as a combination of the ultimatum game with the trust game in the paradigm of trustors' moving first. In Andreoni's game setting, subgame perfect equilibria does not assure efficiency enhancement from satisfaction guarantee unless moral concerns are introduced. By contrast, in our article subgame perfect equilibria predicts improved efficiency from introducing information design.[12]

Our article also contributes to the literature involving information design. In their influential paper, Kamenica and Gentzkow (2011) considers a setting where a designer and a receiver share a common prior about the underlying state and the designer chooses an information structure (a mapping from the underlying state to realized signal) to persuade the receiver to choose an action.[13] Among numerous extensions of the basic setting along various dimensions, one line of extension allows that the two players possess different priors about the underlying state (Hedlund, 2017; Perez-Richet, 2014).[14] Nevertheless, the existing literature about information design still maintains the assumption that both players have exogenous priors about the underlying state. While our setting still shares the feature that the designer chooses an informative structure, it does depart considerably from the existing paradigm: the designer is assumed to (as if) endogenously choose a prior about the underlying state for the receiver. This opens a door for exploring settings where one party chooses both action and information structure to influence the receiver's action.[15] In addition, we provide experimental evidence

[11]See more in Bolton et al. (2004), Charness et al. (2011), Masuda and Nakamura (2012).

[12]Precisely, no trusting act is one of subgame perfect equilibria in Andreoni (2018), but in ours conditional trust, that is, trusting act occurs with positive probability up to one, is of any subgame perfect equilibria.

[13]See Kamenica (2019) and Bergemann and Morris (2019) for surveys in this strand of literature.

[14]See also Alonso and Câmara (2016a,b).

[15]Asriyan et al. (2019) also considers a setting where a designer choose both action and information structure to influence a receiver's action but it deviates the basic setting in more than one aspect: a receiver instead of a designer endogenously chooses an information structure, which

that a considerable fraction of players fail to realize that information structure per se conveys a message about the underlying state, which contrasts with the assumption embedded in existing theories about information design. Recall that our experimental design is a conservative design in the sense that players are explicitly told that information structure is (intentionally) designed by opponents. In reality individuals are not explicitly informed of this, so it is likely that more players in realistic settings fail to understand or incorporate the message conveyed by information structure. Thus, our study suggests that for the sake of building a more realistic theory about information design it seems necessary to model behavioral receivers who are not rather sophisticated, and in particular neglect the message from information structure per se.[16] As a first and small step, we propose a behavioral model with a feature that understanding the message conveyed by information structure is strategically more sophisticated than understanding the message conveyed by signal.

Our article additionally contributes to the literature about informativeness ordering. The classical work by Blackwell (1951, 1953) proposes a statistical definition about informativeness ordering and shows that an information structure is more informative under this definition if and only if it is subjectively more valuable for any Bayesian agent in any decision problem. Since the ordering under this definition is partial, refinements of this definition have been proposed through imposing restrictions on payoff functions and posterior beliefs that apply to specific economic contexts.[17] Informativeness orderings under these refinements are

depends on a few factors including the designer's choice of product complexity; and the designer chooses an action that imposes no restriction on the receiver's prior about the underlying state.

[16]Of course, Kamenica and Gentzkow (2011) makes a point that a sender can still benefit from information design even if "receiver understands that sender chose what information to convey with the intent of manipulating her action for his own benefit". Nevertheless, it is still of theory interest to explore the impact of information design in settings where receiver fails to understand that sender intentionally designs information structure.

[17]See, among others, Lehmann (1988) and Persico (2000) for the criterion of effectiveness, Ganuza and Penalva (2010) for the criterion of super-modular precision and integral precision, and Athey and Levin (2018) for the criterion of monotone information order for non-decreasing objective functions.

still partial and in our two-state setting these proposed orders are equivalent to Blackwell's order.[18] Thus it is impossible for these informative orders to have a numerical representation that maintains one-to-one correspondence.[19] We advance in this direction by introducing and characterizing an index of Blackwell's informativeness that can be interpreted that a larger index value is no less Blackwell informative. In addition, applying these statistical criteria of informativeness ordering to data is typically involved. For example, Blackwell's order includes an existence argument when examining whether one information structure is a garbling of the other. Our article advances by replacing the existence argument with an equivalent condition that are easily checked.

Our article finally contributes to the literature exploring the relationship between prosociality and strategic sophistication. Existing experimental games typically focus on one of the two dimensions exclusively while leaving little room for the other dimension to play a role. Perhaps because of this reason, there are limited studies investigating their relationship.[20] Among existing studies in this regard, measures of strategic sophistication are either directly taken from observed proxies including relevant academic scores or elicited through relevant games involving no prosociality. Although some studies find the two dimensions are not correlated (Arruñada et al., 2015), other studies find they are correlated (Chen et al., 2013; Jones, 2008). Exploring the relationship becomes increasingly important once we consider social contexts in which both prosociality and strategic reasoning conceivably play a crucial role, as in our Game 2. To the best of our knowledge,

[18]According to Ganuza and Penalva (2010), the criterion of integral precision nests other order criteria listed in Footnote 17, all of which nest Blackwell's order. They show in their Theorem 2 that for a binary state the criterion of integral precision is equivalent to Blackwell's order for any interior prior.

[19]Cabrales et al. (2013) refines Blackwell's order by restricting to the special class of ruin-averse utility functions and no-arbitrage investment sets. Their proposed order is complete and is numerically represented by the expected decrease of entropy from the prior to the posteriors.

[20]In a more broad sense, prosociality involves preference side and strategic reasoning concerns belief side, and it is a tradition in the economics profession that preference is treated as orthogonal to belief.

our study is among the first contributions that explore the correlation between the two dimensions in such settings. Baader and Vostroknutov (2017) studies the interaction of the two dimensions through a game of traveller's dilemma and investigates which of the two dimensions accounts for the observed behaviors in their setting. By contrast, our focus in this regard is whether the two dimensions are interdependent and this is a must-answered question when modeling them together. Our finding that they are indeed orthogonal assures that it is plausible to make an independent assumption when modeling the two factors together.

3 Games, Informativeness Ordering, and Equilibrium Analysis

3.1 Two Games

We introduce two experimental games to study whether trustees' being entitled to information design resolves the social dilemma involving trust, and to study the underlying mechanism of success or failure of such intervention. To designate trustees as a leading role in strategic interaction, we flip the order of player move in classical trust games where typically a trustor moves first and a trustee moves second. So a trustee always moves first in our games. Then naturally we use a feature that whether the trustee is provided an option of designing information to differentiate an inverse trust game (Game 1) and a trustworthiness design game (Game 2).

In Game 1, a trustee (player A) with an endowment of 10 moves first to decide the allocation plan on payoff of an investment project and a trustor (player B) with an endowment of 10 moves second to decide whether or not the investment project is carried out without knowing her opponent's choice of allocation plan. If the project is carried out, the endowments are spent and the project generates a total payoff of 30 and the payoff allocation, depending on player A's decision, could be $(15, 15)$ or $(22, 8)$ for the two players. Our pick of the two allocation plans on payoff is to capture a main feature about trust behavior: trust is socially desirable and also beneficial to both parties but placing trust on others is at the risk of being betrayed (See also Bohnet et al. (2008)). Instead of directly choosing either allocation plan, player A decides a probability $p \in [0, 1]$ with which $(15, 15)$ is to be used and correspondingly the remaining probability of using $(22, 8)$. We interpret p as a measure of the level of trustworthiness.[21] Knowing neither the value of p nor the allocation plan that is being used, player B decides $z \in \{1, 0\}$, that is, *invest* or *not invest*. If player B invests in the project then the project is carried out. In this case, player A and player B receive payoff according to the allocation plan that is being used. If player B does not invest in the project then the project is not carried out. In this case, both players keep their initial endowment of 10.

Game 2 introduces a stage of information design into Game 1. In addition to deciding p, a trustee (player A) also decides the conditional likelihoods of signals indicating the allocation plan. Specifically, conditional on the allocation plan that is to be used, either signal $s \in \{b, w\}$ is generated according to a conditional likelihood specified by player A and it becomes observable to player B. Player A decides both $q_1 \in [0, 1]$ and $q_2 \in [0, 1]$, which specify the probability of generating a signal $s = b$ when the allocation plan that is to be used is $(15, 15)$ and $(22, 8)$ respectively. For brevity, we simply use state 1 and state 2 to denote the case when the allocation plan is $(15, 15)$ and $(22, 8)$, and correspondingly call q_1 (q_2) as the conditional likelihood of generating a signal $s = b$ in state 1 (state 2). After observing the conditional likelihoods (q_1, q_2) and a signal generated according to the underlying state and the conditional likelihoods, a trustor (player B) decides $z \in \{1, 0\}$. The payoff rule in Game 2 remains the same as in Game 1.

[21] In classical trust games a trustee moves second and decides different amounts of returning back after receiving investment, and in turn the return amount serves as a measure of the level of trustworthiness. p can also be interpreted alternatively: it could reflect a trustee's determination on choosing $(15, 15)$ or reflect a trustee's applying a mixed strategy to choose payoff allocation.

Discussion. We look at the feature in Game 2 that trustees choose (p, q_1, q_2) and trustors observes (q_1, q_2, s) from two perspectives: its interpretation in our motivating example about financial products design and its usefulness. Consider an investee who makes a marketing plan including a formal investment agreement and informal advertising contents. The nature of the underlying investment, which could turn out to be win-win or fraudulent ex post, is jointly determined by the investee's act and some exogenous shocks including change of market condition. The investee could decide to be trustworthy or fraudulent leaving exogenous shocks for playing no role: choose an act of $p = 1$ in which the nature will be win-win or choose an act of $p = 0$ in which the nature will be fraudulent. Nevertheless, one can imagine that an investee who tend to be fraudulent may allow some exogenous shocks to finally decide the nature so that exogenous shocks can be used as a scapegoat for a fraudulent outcome. An investor's investigation of advertising contents and the formal agreement makes her an impression of either investing recommendation or not investing recommendation, which is the interpretation of a signal s. Suppose that the investee has some influential power on the formation of the investor's either impression by crafting a specific marketing plan. For instance, the marketing plan could be fairly straightforward and using precise language so that the investor forms a different impression when the nature differs. It could also be largely incomprehensible and using misleading language so that the investor forms a similar impression regardless of the nature. Thus, (q_1, q_2) is interpreted as comprehensibility and language use of a marketing plan in this setting.

Regarding the usefulness of the feature, a setting of $p \in [0, 1]$ is a more flexible measure of trustworthiness than a setting of binary $p \in \{0, 1\}$, which is particularly important given that a main variable of interest in this article is trustworthiness. In addition, it helps to generate both an equilibrium with intermediate trustworthiness and an equilibrium with full trustworthiness. A setting of $q_1 \in [0, 1]$ and

$q_2 \in [0,1]$ enables us to seamlessly incorporate existing knowledge of informativeness ordering and information design without any distortion.

3.2 Informativeness ordering

An information structure in Game 2 specifies the conditional likelihood of observing each possible signal given each underlying state. Blackwell (1951, 1953) proposes a definition of ranking two information structures based on a notion of "adding noise". In our setting, an information structure $(q_1, q_2) \equiv Q$ can be represented by 2×2 right stochastic matrix:

$$\begin{pmatrix} q_1 & 1-q_1 \\ q_2 & 1-q_2 \end{pmatrix}$$

We adopt the definition of Blackwell's informativeness order in our setting.[22]

Definition 1 (Blackwell's informativeness order). *An information structure Q is more informative than an information structure Q' ($Q \succsim Q'$) if we can replicate the second information structure from the first one by randomly drawing a signal after each observation of signal under the first one, that is, there exists a right stochastic matrix*

$$\begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix}$$

such that

$$\begin{pmatrix} q_1 & 1-q_1 \\ q_2 & 1-q_2 \end{pmatrix} \begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix} = \begin{pmatrix} q_1^{'} & 1-q_1^{'} \\ q_2^{'} & 1-q_2^{'} \end{pmatrix},$$

where $\alpha, \beta \in [0,1]$. Q is strictly more informative than Q' if $Q \succsim Q'$ but not $Q' \succsim Q$. Q

[22]It is notable that Blackwell shows that the statistic definition is equivalent to an definition based on decision maker's subjective valuation: An information structure Q is more Blackwell informative than an information structure Q' if and only if every Bayesian agent who maximizes his expected utility prefers the first one to the second for any possible decision problem.

and Q' are identically informative if $Q \succsim Q'$ and $Q' \succsim Q$. Q is no less informative than Q' if Q' is not strictly more informative than Q. We say Q and Q' are comparable if either $Q \succsim Q'$ or $Q' \succsim Q$.

The existence argument in Definition 1 makes it inconvenient to use for data analysis. So we derive an easily checked equivalent condition with an intuitive interpretation. We interpret $\frac{q_1}{q_2}$ ($\frac{1-q_1}{1-q_2}$) as the likelihood ratio of generating a black (white) signal in state 1 against in state 2, whose value ranges from 0 to ∞. The equivalent condition simply dictates that the two likelihood ratios under one information structure nests that under another information structure.

Proposition 1. *Q is more informative than Q' if and only if: $[min\{\frac{q_1^{'}}{q_2^{'}}, \frac{1-q_1^{'}}{1-q_2^{'}}\}, max\{\frac{q_1^{'}}{q_2^{'}}, \frac{1-q_1^{'}}{1-q_2^{'}}\}] \subseteq [min\{\frac{q_1}{q_2}, \frac{1-q_1}{1-q_2}\}, max\{\frac{q_1}{q_2}, \frac{1-q_1}{1-q_2}\}]$.*

It is known that Blackwell's order is not complete. In fact, Proposition 1 implies that two information structures are not comparable if and only if there is no nested relationship between the likelihood ratios under the two information structure.

Corollary 1. *Q and Q' are not comparable if and only if: $max\{\frac{q_1}{q_2}, \frac{1-q_1}{1-q_2}\} > max\{\frac{q_1^{'}}{q_2^{'}}, \frac{1-q_1^{'}}{1-q_2^{'}}\}$ and $min\{\frac{q_1}{q_2}, \frac{1-q_1}{1-q_2}\} > min\{\frac{q_1^{'}}{q_2^{'}}, \frac{1-q_1^{'}}{1-q_2^{'}}\}$, or $max\{\frac{q_1}{q_2}, \frac{1-q_1}{1-q_2}\} < max\{\frac{q_1^{'}}{q_2^{'}}, \frac{1-q_1^{'}}{1-q_2^{'}}\}$ and $min\{\frac{q_1}{q_2}, \frac{1-q_1}{1-q_2}\} < min\{\frac{q_1^{'}}{q_2^{'}}, \frac{1-q_1^{'}}{1-q_2^{'}}\}$.*

Proposition 1 also implies that Blackwell's order is transitive. However, the binary relation that two information structures are not comparable is not transitive. For instance, consider three information structures: $(0.3, 0.1)$, $(0.8, 0.6)$, and $(0.2, 0.1)$. While the first two are not comparable and the last two are not comparable, $(0.3, 0.1)$ is more informative than $(0.2, 0.1)$.

Although Blackwell's order is not complete, the most and the least information structure can be cleanly identified. In addition, Blackwell's order is generally asymmetric.

Corollary 2. *Q is the most informative structure, that is, more informative than any other one, if and only if $(q_1, q_2) = (1, 0)/(0, 1)$. Q is the least informative structure if and only if $q_1 = q_2$. Q and Q' are identically informative if and only if $Q + Q' = (1, 1)$ or both are the least informative..*

Since Blackwell's order is not complete, there is no numerical representation in the standard sense, that is, there does not exist a mapping from information structure to real number that satisfies two properties: (1) if Q is strictly more informative than Q' then the former one is assigned a greater number, and if Q and Q' are identically informative than they are assigned a same number; and (2) the converse statement of (1). To appropriately categorize information structure, we consider an index that satisfies only one of the two properties. In our previous example about information structures $(0.3, 0.1)$, $(0.8, 0.6)$, and $(0.2, 0.1)$, the second property requires that $(0.3, 0.1)$ and $(0.8, 0.6)$ have the same index value and that $(0.8, 0.6)$ and $(0.2, 0.1)$ have the same index value. An index that satisfies the second property then has to assign the same index value to $(0.3, 0.1)$ and $(0.2, 0.1)$, which is not in line with the purpose of introducing the index: the index serves to help differentiate the degrees of different information structures' informativeness. So we instead introduce an index that satisfies only the first property. For such index we can interpret that Q is no less informative than Q' if the former one has an index value no less than the second one. In this spirit, we proposes an index that indeed satisfies the first property. [23]

Definition 2. *We define index I_Q for Blackwell's informativeness of Q to be the chance that Q is more informative than any Q' randomly drawn from $[0, 1] \times [0, 1]$ according to the uniform distribution, that is, $I_Q \equiv Pr(Q' \sim [0, 1] \times [0, 1] : Q \text{ is more informative than } Q')$.*

[23] One may consider an alternative definition of the index based on a dual perspective: an index of Q is defined as one minus the chance that any Q' randomly drawn from $[0, 1] \times [0, 1]$ according to the uniform distribution is more informative than Q. It can be established that the alternative index does not satisfy the strict inequality part of the first property, that is, it does not satisfy the requirement that strictly more informative structure has a higher index value.

It turns out that the index can be characterized by the difference between the two likelihoods of an information structure.

Proposition 2. *For any Q, $I_Q = |q_1 - q_2|$.*

Clearly, Propositions 1 and 2 imply that the index I_Q satisfies the first property. In addition, it is clear that $I(q_1, q_2) = 1$ if and only if $(q_1, q_2) = (1, 0)/(0, 1)$ and $I(q_1, q_2) = 0$ if and only if $q_1 = q_2$. Thus, we can strengthen the interpretation of $I_Q = 1/0$: Q is the most (least) informative structure if $I_Q = 1(0)$.

3.3 Equilibrium Analysis

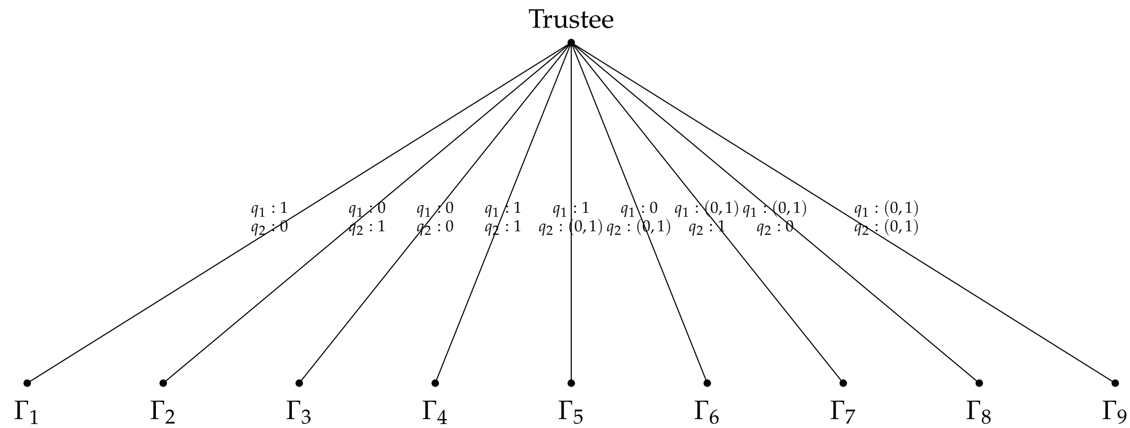
Clearly, Game 1 has a unique Nash equilibrium: a trustee chooses $p = 0$ and a trustor does not invest in the project. In this subsection we explore what actions equilibrium analysis under the standard assumptions prescribes for both players in Game 2.

Definition 3. *In Game 2, a trustee's pure strategy is a three-dimensional action profile: $\sigma_D = (p, q_1, q_2) \in [0, 1] \times [0, 1] \times [0, 1]$. A trustor's pure strategy is a mapping from her information set to her action set: $\sigma_T : [0, 1] \times [0, 1] \times \{B, W\}/\{(1, 1, W)\} \cup \{(0, 0, B)\} \equiv I \mapsto \{invest, not\ invest\} \equiv \{1, 0\}$ with $\sigma_T(q_1, q_2, s) \equiv z \in \{1, 0\}$ for any $(q_1, q_2, s) \in I$.*

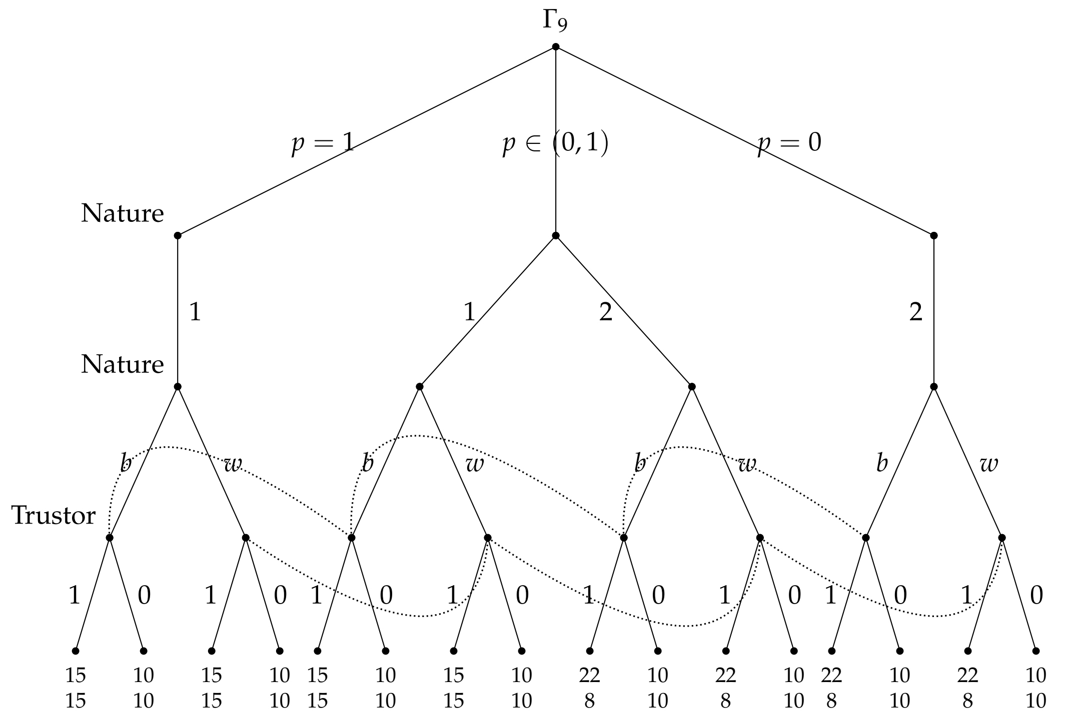
In the spirit of Osborne and Rubinstein (1994) (page 102), we frame and categorize Game 2 into an extensive game with simultaneous moves: a trustee moves first by choosing (q_1, q_2), and then both the trustee and a trustor move simultaneously by choosing p and σ_T respectively. We illustrate the game tree in Figure 1, which demonstrates that there is a proper subgame contingent on each (q_1, q_2).[24] We apply the solution concept of subgame perfect Nash equilibrium (SPNE) in the following equilibrium analysis.[25]

[24]Proper subgames can be classified into nine categories and we illustrate the other eight proper

Figure 1: Game Tree



(a) The entire game



(b) The subgame Γ_9

Definition 4. *In Game 2, denote by* $h \in \{(q_1, q_2, s) : (q_1, q_2, s) \in I\} \cup \{(q_1, q_2) : (q_1, q_2) \in [0,1] \times [0,1]\} \cup \emptyset$ *a nonterminal history in the extensive game* Γ*, denote by*

---

subgames in Appendix A.

[25]We also apply the solution concept of pure strategy Nash equilibrium to Game 2, where no trust equilibria and conditional trust equilibria are charaterized in Appendix A.

$\sigma_D|_h$ and $\sigma_T|_h$ players' strategies that (σ_D, σ_T) induce in the subgame $\Gamma(h)$. A strategy profile $\{\sigma_D, \sigma_T\}$ constitutes a ***pure strategy SPNE*** *if: (1) for every history $h = (q_1, q_2, s)$ the trustor has no incentive to deviate from $\sigma_T|_h$, (2) for every history $h = (q_1, q_2)$ neither the trustee nor the trustor has an incentive to deviate from $(\sigma_D|_h, \sigma_T|_h)$, and (3) for the history $h = \emptyset$ the trustee has no incentive to deviate from σ_D.*

Recall that a typical technique to compute SPNE is replace any proper subgame with assigning the payoff vector associated with a Nash equilibrium in this subgame to the starting node of this subgame. When there are multiple equilibria in the picked subgame, one can choose any of equilibria and by varing the Nash equilibrium for the subgames one can compute all SPNE. Game 2 includes subgames of two levels: the entire game itself and proper subgames dependent on (q_1, q_2). The trustee at the initial node should optimally choose (q_1, q_2) that maximizes his expected payoff in the reduced game. Obviously, the trustee's choosing $p = 1$ and the trutor's choosing *invest* conditional on observing a black ball constitute a Nash equilibrium at the node $(q_1 = 1, q_2 = 0)$ and the trustee's expected payoff is 15 in this equilibrium. The above technique then suggests that the trustee at the initial node should choose (q_1, q_2) that delivers him an expected payoff no less than 15. In fact, we show that the maximal expected payoff the trustee can receive in a Nash equilibrium of any node (q_1, q_2) is 15. Thus, any equilibrium at proper subgames that delivers the trustee an expected payoff less than 15 will not survive from subgame perfection and any equilibrium at proper subgames that delivers the trustee an expected payoff of 15 will survive from subgame perfection. As a result, Game 2 has only one type of pure strategy SPNE that delivers the designer an expected payoff of 15.

Proposition 3. *Game 2 has only one type of pure strategy SPNE,* ***conditional trust equilibria****, which is characterized as:*

$\sigma_D^* = (q_1 = 1, \frac{q_1}{q_2} = \frac{12}{5}, \frac{1}{7} < p < 1)$, $(q_1 = 1, \frac{q_1}{q_2} \geq \frac{12}{5}, p = 1)$, $(1 - q_1 = 1, \frac{1-q_1}{1-q_2} =$

$\frac{12}{5}, \frac{1}{7} < p < 1)$ *or* $(1 - q_1 = 1, \frac{1-q_1}{1-q_2} \geq \frac{12}{5}, p = 1)$;

$$\sigma_T^*(\hat{q_1}, \hat{q_2}, s) = \begin{cases} 1 & \textit{if } (\hat{q_1}, \hat{q_2}, s) \in \{(q_1 = 1, \frac{q_1}{q_2} \geq \frac{12}{5}, b)\} \cup \{(1 - q_1 = 1, \frac{1-q_1}{1-q_2} \geq \frac{12}{5}, w)\}, \\ 0 & \textit{if } (\hat{q_1}, \hat{q_2}, s) \in \{(q_1 = 1, q_2 \leq \frac{12}{5}, w)\} \cup \{(1 - q_1 = 1, 1 - q_2 \leq \frac{12}{5}, b)\}, \\ A_{off} & \textit{if } (\hat{q_1}, \hat{q_2}) \notin \{(q_1 = 1, q_2 \leq \frac{12}{5})\} \cup \{(1 - q_1 = 1, 1 - q_2 \leq \frac{12}{5})\}. \end{cases}$$

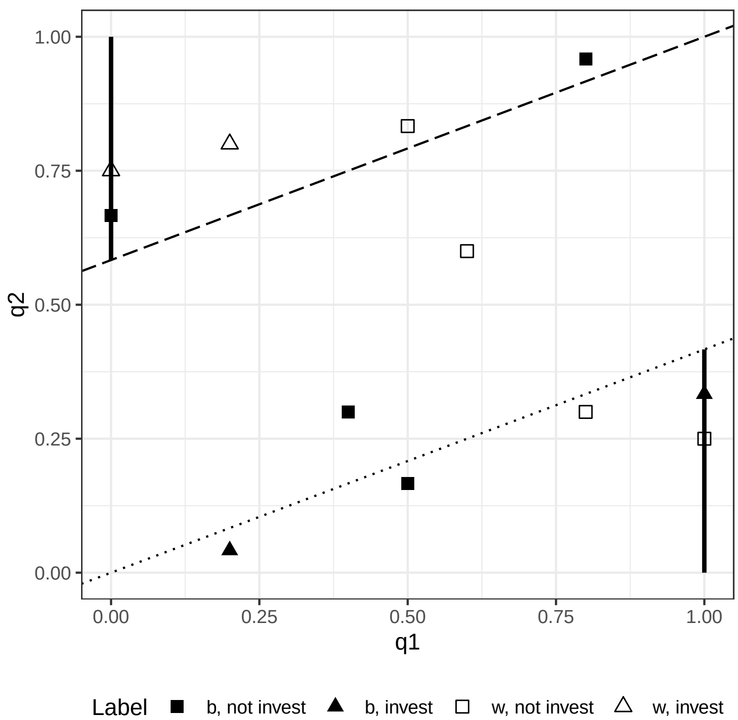
Actions on the off-equilibrium path can be specified as:

$$A_{off} = \begin{cases} 1/0 & \textit{if } (\frac{\hat{q_1}}{\hat{q_2}} \geq \frac{12}{5}, s = b) \textit{ or } (\frac{1-\hat{q_1}}{1-\hat{q_2}} \geq \frac{12}{5}, s = w), \\ 0 & \textit{otherwise.} \end{cases}$$

We illustrate the equilibrium strategy in Game 2 for trustors in Figure 2.

Remark. It is notable that among all pure strategy SPNE, some equilibria seem more plausible than others. For example, while both $(q_1 = 1, q_2 = \frac{1}{3}, p = 1)$ and $(q_1 = 1, q_2 = 0, p = 1)$ are the trustee's equilibrium strategies one could argue that it is more plausible for the trustee to choose $(q_1 = 1, q_2 = 0)$ over $(q_1 = 1, q_2 = \frac{1}{3})$. Specifically, at the node $(q_1 = 1, q_2 = \frac{1}{3})$ the trustee's choosing $p = 1$ and the trustor's choosing *invest* conditional on observing a black signal constitute only one of two Nash equilibria. Indeed there is also another Nash equilibrium at this node: the trustee chooses $p = 0$ and the trustor always chooses *not invest*, where the trustee receives an expected payoff of 10. At the node $(q_1 = 1, q_2 = 0)$ the trustee's choosing $p = 1$ and the trustor's choosing *invest* conditional on observing a black signal constitute the unique Nash equilibrium. Thus, choosing $(q_1 = 1, q_2 = 0)$ at the initial node guarantees that the trustee arrives at a Nash equilibrium with an expected payoff of 15 in the subsequent node but choosing $(q_1 = 1, q_2 = \frac{1}{3})$ at the initial node may make the trustee arrive at either a Nash equilibrium with an expected payoff of 15 or a Nash equilibrium with an expected payoff of 10 in the subsequent node. So one could argue that the trustee is likely

FIGURE 2: Trustor's equilibrium strategy in Game 2



Notes: The dotted line is $\frac{q_1}{q_2} = \frac{12}{5}$ and the dashed line is $\frac{1-q_1}{1-q_2} = \frac{12}{5}$.

to be inclined to choosing $(q_1 = 1, q_2 = 0)$ over $(q_1 = 1, q_2 = \frac{1}{3})$. This observation helps to exclude all the trustee's SPNE strategies but $(q_1 = 1, q_2 = 0, p = 1)$ and $(q_1 = 0, q_2 = 1, p = 1)$. In other words, the most refined equilibrium seems to be that the trustee chooses to be fully trustworthy with the most informative structure. One could formalize the notion of perfection and further refine the solution concept of pure strategy SPNE. Instead we decide to leave it be answered empirically in the latter section.

Finally, we compare equilibrium outcomes of the two games in the following three aspects. First, the trustee chooses to be trustworthy with positive probability up to full probability in Game 2, that is, $p \in (\frac{1}{7}, 1]$, while he chooses to be trustworthy with zero probability in Game 1. Second, while the trustor does not invest in Game 1, in Game 2 she invests on the equilibrium paths $(q_1 =$

$1, \frac{q_1}{q_2} > \frac{12}{5})$ and $(1 - q_1 = 1, \frac{1-q_1}{1-q_2} > \frac{12}{5})$ and invests on the equilibrium paths $(q_1 = 1, q_2 = \frac{5}{12})$ and $(1 - q_1 = 1, 1 - q_2 = \frac{5}{12})$ with a probability exceeding $\frac{1}{2}$, that is, $p + (1-p)q_2 = p + (1-p) * \frac{5}{12} = \frac{5}{12} + \frac{7}{12}p > \frac{5}{12} + \frac{7}{12} * \frac{1}{7} = \frac{1}{2}$. Third, in terms of expected payoff ex ante, the payoff in Game 1 is 10 for both players and in Game 2 the payoff becomes 15 for the trustee and exceeds 10 for the trustor.[26]

4 Experimental Design and Procedure

4.1 Experimental Design

We design a theory-guided experiment with two main objectives. First, we examine whether introducing information design about trustworthiness causes more socially desirable outcomes, that is, increase the level of trustworthiness and the fraction of trusting act. Second, we would like to understand the underlying reasons and investigate in Game 2 how trustees associate choice of trustworthiness with choice of information structure, how trustors respond to potentially different sets of information structure and signal, and to what extent their actions are consistent with what are prescribed by the theory.

We employ a within-subject design according to which each subject plays both Game 1 and Game 2. Each subject is randomly assigned a constant player role through the experiment, either player A or player B, and is paired with an anonymously new opponent in each play.[27] Both games are described to subjects that player B decides whether an investment project is carried out and player A decides the chances that payoff allocations $(15, 15)$ and $(22, 8)$ are used in the case of the project being carried out. The task of deciding the chance of p is described

[26]For the trustor, $15p + (1-p)[8q_2 + 10(1-q_2)] = 15$ or $= 15p + (1-p)[8 * \frac{5}{12} + 10(1 - \frac{5}{12})] = \frac{110}{12} + \frac{70}{12}p > \frac{110}{12} + \frac{70}{12} * \frac{1}{7} > 10$.

[27]Burks et al. (2003) provides experimental evidence that a subject's playing both roles brings about confounds and is likely to reduce the degree of trust and reciprocity in the trust game. In addition, given our within-subject design, asking a subject to play each role of each game is likely to cause confusion.

to subjects in the following intuitive way: (1) pick an integer number P from 0 to 100; (2) if $P = 100$ then the payoff allocation $(15, 15)$ is used and if $P = 0$ then the payoff allocation $(22, 8)$ is used; and (3) if $0 < P < 100$ then the payoff allocation $(15, 15)$ is used with a chance of P% and the payoff allocation $(22, 8)$ is used with a chance of $(100 - P)$%. Similarly, the task of deciding the conditional likelihood q_i $(i = 1, 2)$ is described to subjects in the following way: (1) when the payoff allocation $(15, 15)$ $((22, 8))$ is used a ball from urn 1 (urn 2) with 100 balls in total is randomly drawn; (2) pick an integer number Q_i from 0 to 100 so that urn i contains Q_i black balls and $100 - Q_i$ white balls.

The decision timing of Game 1 goes as follows. Player A first decides a number P. Then player B decides whether to invest in the project without knowing player A's choice of P and which of the two payoff allocation is used. The decision timing of Game 2 proceeds as follows. Player A first decides the numbers P, Q_1, Q_2 . Then a payoff allocation is chosen to be used by a computer server according to the number P and a ball is randomly drawn from the corresponding urn. Finally, player B observes only the numbers Q_1, Q_2 and the color of the randomly drawn ball, and decides whether to invest in the project.

Each player plays five rounds of Game 1 and five rounds of Game 2, and the order of the two games, that is, five rounds of Game 1 followed or preceded by five rounds of Game 2, is varied across sessions to control game order effect.[28] Each player is paired with an anonymously new opponent in each of the ten rounds. In addition, the outcome of each round is not revealed to subjects. These two settings are to exclude or suppress the confounds that includes reputation concern, learning about opponents and the effect of past experience. The payoff unit in this experiment is token and subjects are paid according to their earned tokens in one round that is randomly selected out of ten rounds so that they have no

[28]Recall that player B's strategy in Game 2 specifies her action in each possible information set. Playing multiple rounds of Game 2 reveals more footprints of player B and enhances the degree of identifying her strategy. To make the number of observations in both games comparable, the number of rounds played for Game 1 is set to the same as for Game 2.

incentive to play different strategies across games to hedge against risks. Each token is redeemed for two Chinese yuan. In addition to the earned payoff in the games, they also receive a show-up fee of five Chinese yuan. Finally, subjects are asked to complete a six-question survey after the paid rounds end.[29]

To make sure subjects understand the two games, we ask subjects to practice playing games before the paid rounds start. Meanwhile, we want to make sure that practice experience does not affect subjects' beliefs and plays in the paid experiment. To this end, we ask each subject, by playing both role A and role B, to play against a computer opponent which always makes a random decision in the practice part. Specifically, each subject plays four rounds in total for practice as a role of player A in Game 1, player A in Game 2, player B in Game 1 and player B in Game 2. The computer opponent is set to make a random decision when it determines P, (P, Q_1, Q_2), and whether to invest in the project. Subjects are informed of the computer opponent's strategy, and they can observe the allocation plan that is used and their payoffs at the end of each practice round.

4.2 Experimental Procedure

There were twenty participants in each session of our experiment. We randomly designated ten of them as player A and the other ten as player B. Each participant was provided with a hard copy of experimental instructions. When all participants were seated and assigned with identification numbers, a pre-recorded audio of experimental instructions was played through louder speakers so as to maintain uniformity across sessions. After the audio was played, an experimental investiga-

[29]The six questions are: (1) report the minimum acceptable value of X for the subject to choose a lottery that has a payoff of 15 with the chance X% and has a payoff of 8 with the remaining chance over a deterministic payoff of 10; (2) report the minimum acceptable value of Y for the subject to choose a lottery that has a payoff of 22 with the chance Y% and has a payoff of 8 with the remaining chance over a deterministic payoff of 15; (3) whether player A on average receives a higher/lower/identical payoff in Game 2 compared to in Game 1; (4) whether player B on average receives a higher/lower/identical payoff in Game 2 compared to in Game 1; (5) whether the subject has taken courses about probability or statistics; and (6) the subject's gender.

tor was responsive to any questions about the understanding of the experimental instructions. The experiment started when participants' questions were all addressed. The experimental interface was made through z-Tree (Fischbacher, 2007), and it ran for four practice rounds, ten paid rounds and then a post-experiment survey. At the end of the experiment, each subject's earnings were displayed only on her computer screen, and conditional on being called upon her participation identification number the subject was paid privately and then left the laboratory. Each session on average lasted for 35 minutes.

We conducted sixteen sessions of the experiment at the Finance and Economics Experimental Economics Laboratory (FEEL) of Xiamen University, with five sessions on December 6th, eight sessions on December 7th and three sessions on December 8th of 2019. Game 2 preceded Game 1 in the eight sessions on December 7th and Game 1 preceded Game 2 in the other eight sessions. A total of 320 subjects in these sixteen sessions were recruited from students at Xiamen University, each of whom participated in only one session. Subjects playing role A and playing role B on average earned 32.6 and 26.05 Chinese yuan (approximately 4.66 and 3.72 US dollars) respectively, compared to a local minimum hourly wage of 18 Chinese yuan in that year.

5 Experimental Results

An observation in our sample refers to the values of a set of the following variables. Main variables of interest (when applicable) include trustworthiness measured by $p = \frac{P}{100}$, information structure measured by $(q_1 = \frac{Q_1}{100}, q_2 = \frac{Q_2}{100})$, signal indicated by the color of a drawn ball ($s \in \{b, w\}$), trust indicated by the action of *invest* or *not invest* ($z \in \{1, 0\}$), and a dummy indicating whether the observation is from Game 1 or from Game 2. Variables of background information include subject identification number, session identification number, dummies (indicating player role, indicating whether Game 1 precedes Game 2, and indicating the allocation plan that is used), round number indexed from 1 to 5, and the payoff of the subject in this round. In addition, the values of subject specific variables collected in the survey are appended to each observation. Overall, the experiment collected 3200 observations, among which one half are from 160 trustee subjects and the other half are from 160 trustor subjects.

We report descriptive statistics of main variables of interest in Table 1. We find few round effects but considerable game order effect about actions in Game 1. Game order effect suggests that we need to take this into account when we analyze treatment effect in next subsection.

Table 1: Descriptive Statistics

Variables	Mean (Std)	Median	Game order effect	Round effect
p^1	0.232 (0.357)	0	0.0119^{**}	$1/4^{**}, 1/5^{***}, 3/5^{***}, 2/5^{**}$
z^1	0.336 (0.473)	0	0.001^{***}	no
p^2	0.435 (0.436)	0.3	0.2474	no
q_1	0.636 (0.303)	0.6	0.2634	no
q_2	0.402 (0.308)	0.5	0.5451	no
z^2	0.47 (0.499)	0	0.288	no

Note: p^i, z^i ($i = 1/2$) refers to the values of p and z in game i, where $z = 1/0$ refers to *invest/not invest* respectively. Column 4 records the p-values of tests (two-sided t-test for p^1, p^2, q_1 and q_2; two-sided proportion test for z^1 and z^2) on the hypothesis that the difference is zero in mean/proportion between the subsample where Game 1 precedes Game 2 and the subsample where Game 1 follows Game 2. Column 5 reports the p-values of tests (two-sided paired t-test for p^1, p^2, q_1 and q_2; two-sided proportion test for z^1 and z^2) on the pair-wise equality of mean values between the corresponding rounds. $***/**$ denote significance at the one/five percent level.

5.1 Results (in)consistent with equilibrium analysis

Table 2 shows that the treatment effect (the change from Game 1 to Game 2) is generally consistent with equilibrium analysis in terms of direction but not magnitude. First, the mean value of trustees' chosen chance of plan $(15, 15)$ being used, p, increases by 87.5% from 0.232 to 0.435. Statistically, the mean of p in Game

2 is significantly higher than that in Game 1 at the one percent level ($p < 0.001$, one-sided Welch t-test and one-sided Wilcoxon rank sum test). Additionally, statistical tests on the equality of the distributions of p in the two games reject the null hypothesis at the one percent level ($p < 0.001$, two-sided Kolmogorov-Smirnov test and two-sided Wilcoxon rank sum test). Second, the fraction of rounds in which trustors choose to *invest* increases by 39.9% from 0.336 to 0.470. A two-sample proportion test shows that the fraction in Game 2 is statistically higher than in Game 1 at the one percent level ($p < 0.001$, one-sided). Third, while we do not find there is significant difference in the average payoff of trustees between the two games, we find that the average payoff of trustors increases significantly from Game 1 to Game 2.[30] Interestingly, among the three belief options of being higher, lower and equal in the post-experiment survey over half of the subjects believe that trustees receive a higher payoff in Game 2 and over three quarters of the subjects believe that trustors receive a higher payoff in Game 2.

When looking at each role in each game, we find that many of subjects' actions deviate considerably from equilibrium predictions although some are aligned with the predictions. In Game 1, the percentages of rounds in which trustees choose $p = 0$, $0 < p \leq 0.5$, $0.5 < p < 1$, and $p = 1$ are 57.75%, 22.25%, 9.5% and 10.5% respectively; and the percentages of rounds in which trustors choose to *invest* and *not invest* are 23.2% and 76.8% respectively. In Game 2, the equilibrium predictions on trustees' actions are $(q_1 = 1, q_2 < \frac{5}{12})$ or $(1 - q_1 = 1, 1 - q_2 < \frac{5}{12})$

[30]The insignificant effect of trustees' payoff change across games is possibly due to the specific experimental parameter setting that the increment in trustees' payoff from no trusting act to trusting act under the plan $(15, 15)$ is less than half of the loss in their payoff from trusting act to no trusting act under the plan $(22, 8)$, that is, $15 - 10 < \frac{1}{2}(22 - 10)$. This can be illustrated from a decomposition of payoff changes. The numbers of plan $(15, 15)$ used in Game 1 and Game 2 are 191 and 351 respectively. Among these rounds, the numbers of trusting act are 67 and 262 respectively (fraction: 35.1% and 74.6%). The numbers of plan $(22, 8)$ used in Game 1 and Game 2 are 609 and 449 respectively. Among these rounds, the numbers of trusting act are 202 and 114 respectively (fraction: 33.2% and 25.4%). Thus, the (relative) increment of the sum of trustees' payoff changes across games is 975 ($= (262 - 67) * (15 - 10)$) and the (relative) loss is 1056 ($= (202 - 114) * (22 - 10)$), which offset with each other although the increase in trusting act across games when the plan is $(15, 15)$ is more than twice as much as the decrease in trusting act across games when the plan is $(22, 8)$.

Table 2: Increase in trustworthiness and trust

	Game 1	Game 2	p-value
trustworthiness (p)	0.232	0.435	2.2×10^{-16}
	0	$> \frac{1}{7}$	
	0.264	0.417	3.4×10^{-8}
	0.2	0.453	2.2×10^{-16}
trust (z)	0.336	0.47	3.3×10^{-8}
	0	$> \frac{1}{2}$	
	0.393	0.45	0.058
	0.28	0.49	8.2×10^{-10}
trustee's payoff	13.449	13.348	0.6668
	10	15	
	13.94	13.32	0.966
	12.958	13.378	0.0962
trustor's payoff	9.914	11.353	2.2×10^{-16}
	10	> 10	
	9.985	11.183	2.2×10^{-13}
	9.843	11.523	2.2×10^{-16}

Note: The corresponding rows 1-4 report the value of the full sample, equilibrium prediction, and the values of subsample 1 and subsample 2 in which Game 1 is firstly and secondly played respectively. One-sided proportional test for variable z and one-sided Welch t-test for the other three variables.

with the corresponding $p = 1$, or $(q_1 = 1, q_2 = \frac{5}{12})$ or $(1 - q_1 = 1, 1 - q_2 = \frac{5}{12})$ with the corresponding $p \in (\frac{1}{7}, 1]$. We find that 24.25% of observations ($\frac{194}{800}$) are aligned with the predictions. All of them belong to the following categories: $(p = 1, q_1 = 1, q_2 = 0)$ (163 observations) and $(p = 1, q_1 = 0, q_2 = 1)$ (31 observations), which echoes our discussion about further equilibrium refinement in Section 3.3. We also consider allowing for 10% perturbation of equilibrium predictions, that is, the predicted actions are $(q_1 \geq 0.9, q_2 < \frac{5*1.1}{12})$ or $(1 - q_1 \geq 0.9, 1 - q_2 < \frac{5*1.1}{12})$ with the corresponding $p \geq 0.9$, or $(q_1 \geq 0.9, \frac{0.9*5}{12} \leq q_2 \leq \frac{1.1*5}{12})$ or $(1 - q_1 \geq 0.9, \frac{0.9*5}{12} \leq 1 - q_2 \leq \frac{1.1*5}{12})$ with the corresponding $p \in (\frac{0.9}{7}, 1]$. When employing the modified criterion for checking consistency, there is only slight improvement: 25.875% of observations ($\frac{207}{800}$) are aligned with the predictions. Depending on whether observed information set is on equilibrium path or off-equilibrium path and also the predicted action, Table 3 reports the consistency of trustors' actions with equilibrium predictions in Game 2. Since trustees choose the most informative structure when they are on equilibrium path, it is not surprising that the consistency ratio of trustors' actions exceeds 90% on equilibrium path. It is noteworthy that when the predicted action is unequivocal on off-equilibrium path trustors' actions are not consistent with the prediction about 27% of the time.

Table 3: Consistency of trustors' actions in Game 2

Information set	Predicted action (z)	Observations	Consistent observations (Ratio)
$(q_1 = 1, \frac{q_1}{q_2} \geq \frac{12}{5}, b)$ $(1 - q_1 = 1, \frac{1-q_1}{1-q_2} \geq \frac{12}{5}, w)$	1	201	195(97.01%)
$(q_1 = 1, \frac{q_1}{q_2} \geq \frac{12}{5}, w)$ $(1 - q_1 = 1, \frac{1-q_1}{1-q_2} \geq \frac{12}{5}, b)$	0	12	11(91.67%)
$(q_1 < 1, \frac{q_1}{q_2} \geq \frac{12}{5}, b)$ $(1 - q_1 < 1, \frac{1-q_1}{1-q_2} \geq \frac{12}{5}, w)$	1/0	41	41(100%)
$(q_1 < 1, \frac{q_1}{q_2} \geq \frac{12}{5}, w)$ $(1 - q_1 < 1, \frac{1-q_1}{1-q_2} \geq \frac{12}{5}, b)$ $(max(\frac{q_1}{q_2}, \frac{1-q_1}{1-q_2}) < \frac{12}{5}, b/w)$	0	546	401(73.44%)
equilibrium path	–	213	206(96.71%)
off-equilibrium path	–	587	442(75.30%)
overall	–	800	648(81%)

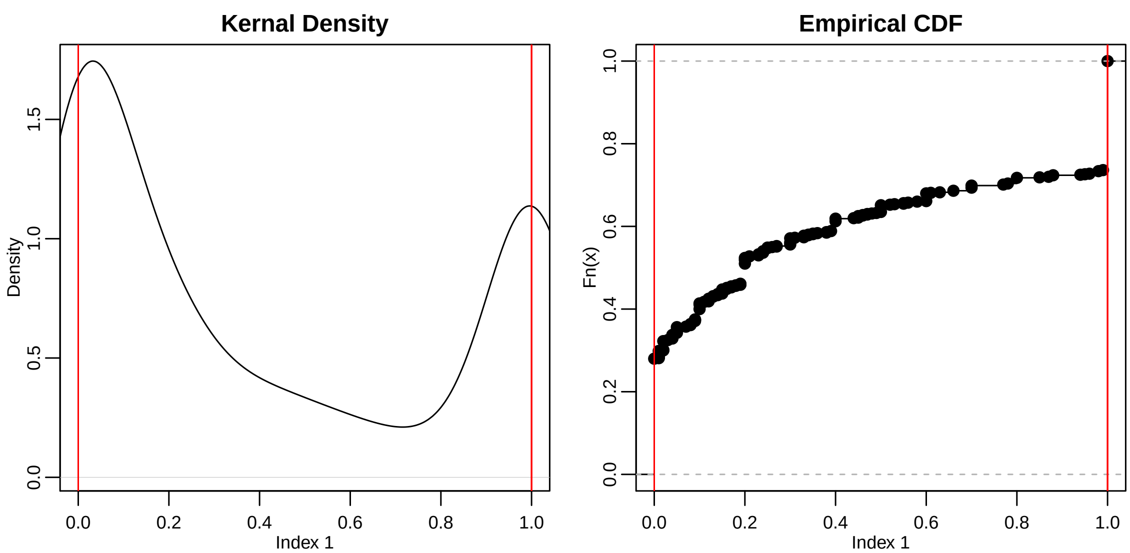
5.2 Trustees' behavioral patterns

We explore trustees' major behavioral patterns from three perspectives below: choice of information structure at the aggregate level; and at the individual level how a trustee's choice of p in Game 2 is associated with his choice of p in Game 1 and how his choice of information structure (q_1, q_2) is associated with his choice of p in Game 2.

Figure 3 illustrates our data sample's empirical distribution of Blackwell's informativeness index I_Q. More precisely, the fraction of $I_Q = 0$ is 28% (224 out of 800), the fraction of $I_Q \leq 0.5$ is 65.125% (521 out of 800), and the fraction of $I_Q = 1$ is 26.375% (211 out of 800). This shows that more than one quarter of information structures are the most informative and more than one quarter of information structures are the least informative.

Pattern 1 (Information structure) *More than half of information structures trustees choose are the most or the least informative, with the most being about* 26% *and the least being* 28%*.*

FIGURE 3: Empirical Distribution of Informativeness Index



Notes: The quartiles of I_Q are 0, 0.2, 1 and 1, with a mean value of 0.3993.

Let $\bar{p}$ be the mean value of p that is chosen by a trustee in five rounds of a specific game. We simply divide $\bar{p}$ into two categories: low $\bar{p}$ if $\bar{p} \leq 0.5$ and high $\bar{p}$ if $\bar{p} > 0.5$. We report in Table 4 the fraction of trustees who have low $\bar{p}$ in both games, the fraction of trustees who have high $\bar{p}$ in both games, the fraction of trustees who have low $\bar{p}$ in Game 1 and have high $\bar{p}$ in Game 2, and the fraction of trustees who have high $\bar{p}$ in Game 1 and have low $\bar{p}$ in Game 2. Table 4 clearly shows that most of trustees do not choose a lower p in Game 2 than in Game 1, the fraction is 97.5%. In other words, most of trustees choose to become on average more trustworthy when the option of information design is introduced.

Table 4: The Association of $\bar{p}$ in Games 1 and 2

	low $\bar{p}$ in Game 2	high $\bar{p}$ in Game 2	total
low $\bar{p}$ in Game 1	55.625%(89)	28.125%(45)	83.75%(134)
high $\bar{p}$ in Game 1	2.5%(4)	13.75%(22)	16.25%(26)
total	58.125%(93)	41.875%(67)	

Note: The percentage refers to the fraction of trustees whose choice of p exhibits the specified pattern among all 160 trustees. The corresponding number of trustees is reported in the parenthesis.

Pattern 2 (Change of trustworthiness) *Most of trustees' mean values of p in Game* 2 *is no less than that in Game* 1*.*

We now investigate how a trustee's choice of p is associated with his choice of information structure Q. Specifically, we investigate at the individual level the extent to which a trustee's choosing informativeness of Q in five rounds is associated with his $\bar{p}$ in Game 2, that is, whether a trustee with a low or high $\bar{p}$ chooses a more informative structure.[31] For a trustee subject, we observe five information structures he chose in five rounds and we divide individual's choice of information structure into three categories: we say he chooses low informativeness if none or only one of his five information structures belongs to information structure of low index ($I_Q \leq 0.5$); we say he chooses high informativeness if four or all of his five information structures belong to information structure of high index ($I_Q > 0.5$); and we say he chooses random informativeness if two or three

[31] Alternatively, one can investigate the association at the round level, which is akin to a perspective at the aggregate level and through which the message is similar to our investigation at the individual level. By defining low/high p and low/high index of information structure similarly, the fractions of rounds are respectively: (low p, low index), 54.375%; (low p, high index), 6% ; (high p, low index),10.75% ; and (high p, high index), 28.875% . Particularly, the fractions of $(p = 0, q_1 = q_2)$ and $(p = 1, |q_1 - q_2| = 1)$ are 16% and 24.25%.

of his five information structures belong to information structure of low index and the remaining belongs to information structure of high index. Table 5 reports the fraction of trustees with high or low $\bar{p}$ who choose low informativeness, high informativeness, or random informativeness. It suggests that some trustees signal their high trustworthiness by choosing high informativeness and other trustees obscure their low trustworthiness by choosing low informativeness.

TABLE 5: The association of $\bar{p}$ and choice of informativeness in Game 2

	low informativeness	high informativeness	random informativeness	total
low $\bar{p}$	45%(72)	2.5%(4)	10.625%(17)	58.125%(93)
high $\bar{p}$	8.75%(14)	20.625%(33)	12.5%(20)	41.875%(67)
total	53.75%(86)	23.125%(37)	23.125%(37)	

Note: Low informativeness indicates $\#(I_Q \leq 0.5) = 4$ or 5, high informativeness indicates $\#(I_Q > 0.5) = 4$ or 5, and random informativeness indicates $\#(I_Q \leq 0.5) = 2$ or 3. The number of subjects in each category is reported in the parenthesis.

Pattern 3 (Association of trustworthiness and information structure) *Among those subjects who choose a low p on average, the majority* ($77.4\% = \frac{72}{93}$) *choose low informativeness and a considerable fraction of them* ($18.3\% = \frac{17}{93}$) *choose random informativeness. Among those who choose a high p on average, the majority choose either high informativeness or random informativeness* ($79.1\% = \frac{53}{67}$)*. Few of them choose both low p on average and high informativeness or choose both high p on average and low informativeness.*

5.3 Trustors' behavioral patterns

We investigate trustor's action pattern contingent on information structure and signal both at the aggregate level and at the individual level. In addition, we investigate how trustor's strategy in Game 2 is related to her strategy in Game 1.

We categorize information set based on a dichotomy of information structure and a dichotomy of signal generated from each specific information structure. As in 5.2, we divide information structure into two categories: information structure of low index ($I_Q \leq 0.5$) and information structure of high index ($I_Q > 0.5$). Since our interest is when the trustor chooses to *invest*, we categorize a signal generated from an information structure, $s|Q$, based on whether it recommends *invest* or it does not recommend *invest*. Note that a trustor's optimal state-dependent action is *invest* if and only if the underlying state is state 1. So we say $s|Q$ recommends *invest* if Q is not the least informative and the likelihood of generating the signal in state 1 is higher than in state 2, that is, $(\frac{q_1}{q_2} > 1, s = b)$ or $(\frac{q_1}{q_2} < 1, s = w)$. Correspondingly, we say $s|Q$ does not recommend *invest* if Q is the least informative or if it is not the least informative and the likelihood of generating the signal in state 1 is lower than in state 2. Overall, there are four categories of information sets: (1) H_1, information structure of low index and signal not recommending *invest*, (2) H_2, information structure of high index and signal not recommending *invest*, (3) H_3, information structure of low index and signal recommending *invest*, and (4) H_4, information structure of high index and signal recommending *invest*.

Table 6 reports the number of each category of information sets and the corresponding fraction of *invest* action. The main message is that at the aggregate level the frequency of subjects' choosing to *invest* is contingent on categories to which the information set belongs. As expected, trustor is very likely to *invest* when observing information structure of high index and signal recommending *invest* and trustor is very unlikely to *invest* when observing information structure of high index and signal not recommending *invest*. An interesting observation is the frequency of *invest* is considerable (about 45%) when observing information structure of low index and signal recommending *invest*. Given the fact that information structure of low index is associated with a low level of trustworthiness, sophisticated trustors would better choose *not invest* when observing information structure of low index. Nevertheless, this observation indicates that some trustor subjects may fail to incorporate the message conveyed through information structure.

Table 6: Information sets and fractions of *invest* in Game 2

Observed information set	H_1	H_2	H_3	H_4
observation #	364	49	157	230
invest (#)	21.70%(79)	10.20%(5)	45.22%(71)	96.09%(221)

Note: $H_1 \equiv \{I_Q \in (0, 0.5], (\frac{q_1}{q_2} > 1, s = w)/(\frac{q_1}{q_2} < 1, s = b)\} \cup \{I_Q = 0, s = b/w\}, H_2 \equiv \{I_Q > 0.5, (\frac{q_1}{q_2} > 1, s = w)/(\frac{q_1}{q_2} < 1, s = b)\}, H_3 \equiv \{I_Q \in (0, 0.5], (\frac{q_1}{q_2} > 1, s = b)/(\frac{q_1}{q_2} < 1, s = w)\}, H_4 \equiv \{I_Q > 0.5, (\frac{q_1}{q_2} > 1, s = b)/(\frac{q_1}{q_2} < 1, s = w)\}$. The fractions of *invest* action in the subsamples of $\{I_Q = 0, s = b/w\}$ and of $\{I_Q \in (0, 0.5], (\frac{q_1}{q_2} > 1, s = w)/(\frac{q_1}{q_2} < 1, s = b)\}$ are 21.88%(49/224) and 21.43%(30/140) respectively.

Pattern 4 (Actions contingent on information set) *The frequency of choosing to invest is the highest (about* 96%*) when observing information structure of high index and signal recommending invest; and the frequency is the lowest (about* 10%*) when observing information structure of high index and not signal not recommending invest. There is also a considerable frequency of choosing to invest (about* 45%*) when observing information structure of low index and signal recommending invest.*

An investigation of information set dependent actions at the individual level involves the distribution of trustor's strategy, which reports the number of subjects who choose each specific strategy. Given our classification of information sets into four categories, trustor's strategy space consists of 16 strategies, each of which specifies a binary action of *invest/not invest* for each category of information sets. Three considerations suggest that an exploration of such distribution provides not much information. First, a strategy space consisting of 16 strategies is still much scattered given a sample of 160 trustor subjects, which leads to a small sample size of subjects who choose each specific strategy. Second, a subject may only observe some of four categories of information sets in five rounds and in turn she can be classified as multiple strategy types. For example, a subject who observes only two (one) categories of information sets will be classified as four (eight) strategy types. Third, when a subject observes a specific category of information set in more than one round, she may choose *invest* in one round and choose *not invest* in another round and consequently we cannot classify her as any of the 16 strategy types. In particular, it is very likely to be the case when observing information set of $(q_1, q_2 = q_1, b/w)$. As expected, our exploration of such strategy distribution indeed turns out not to be informative.

Taking into account the three considerations, we instead look at the distribution coarsely: investigate trustor's action for each category of information set, H_i, in the subsample of trustors whose action in that specific H_i is self-consistent. Take H_1 for example, we look at the subsample in which a subject takes the same action, *invest* or *not invest*, in all the rounds he/she observes H_1. Overall, 159 out of 160 subjects take self-consistent action in at least one H_i. Among them, 117, 44, 100 and 121 subjects take self-consistent action in information sets H_1, H_2, H_3 and H_4 respectively. Additionally, for each category of information set, we investigate the distribution of trustor's action in a subsample in which a trustor takes self-consistent actions in *any* H_i he/she observes. Overall, 103 out of 160 subjects belong to this subsample. Among them, 101, 27, 73 and 85 subjects choose self-consistent action in information sets H_1, H_2, H_3 and H_4 respectively.

Table 7 reports the distribution of trustor's action in each category of information sets. These findings at individual level is similar to Pattern 4, which is at the aggregate level. Both suggest that trustor subjects are able to make inference about the underlying state based on the given information structure and its generated signal, but some of them fail to incorporate the message embedded in the information structure per se.

Pattern 5 (Distribution of trustor's strategy) *The fraction of trustor subjects who choose invest is the highest (about* 98%*) when observing information structure of high index and signal recommending invest. The fraction is the smallest (about* 10%*) when observing information structure of high index and signal not recommending invest. In addition, a considerable fraction of subjects (about* 43%*) choose invest when observing*

Table 7: Distribution of trustor's action in Game 2

	H_1	H_2	H_3	H_4
Action in *an* H_i does not vary	$10.26\%(\frac{12}{117})$	$9.09\%(\frac{4}{44})$	$43\%(\frac{43}{100})$	$98.35\%(\frac{119}{121})$
Action in *each* H_i does not vary	$10.89\%(\frac{11}{101})$	$11.11\%(\frac{3}{27})$	$39.73\%(\frac{29}{73})$	$98.82\%(\frac{84}{85})$

Note: $H_1 \equiv \{I_Q \in (0, 0.5], (\frac{q_1}{q_2} > 1, s = w)/(\frac{q_1}{q_2} < 1, s = b)\} \cup \{I_Q = 0, s = b/w\}, H_2 \equiv \{I_Q > 0.5, (\frac{q_1}{q_2} > 1, s = w)/(\frac{q_1}{q_2} < 1, s = b)\}, H_3 \equiv \{I_Q \in (0, 0.5], (\frac{q_1}{q_2} > 1, s = b)/(\frac{q_1}{q_2} < 1, s = w)\}, H_4 \equiv \{I_Q > 0.5, (\frac{q_1}{q_2} > 1, s = b)/(\frac{q_1}{q_2} < 1, s = w)\}$. The denominator and the numerator denote the number of subjects in a subsample and the number of those who choose *invest* respectively.

information structure of low index and signal recommending invest.

We finally investigate the association of trustor's strategy in the two games. In Game 1, a trustor subject is labeled as *trust* if she invests in three or over three rounds and is labeled as *no trust* otherwise. Based on this classification, there are 50 *trust* subjects and 110 *no trust* subjects. Since information sets that a subject observes in Game 2 are determined by her opponent and it has nothing to do with her classification, it is expected that the distribution of four categories of information sets that subjects observe in Game 2 is similar for *trust* subjects and *no trust* subjects in Game 1, which is indeed the case according to Table 8.

Table 8: Distribution of information sets for *trust/no trust* subjects of Game 1

		H_1	H_2	H_3	H_4
trust subjects of Game 1	count	123	12	51	64
	expected	113.75	15.31	49.06	71.87
no trust subjects of Game 1	count	241	37	106	166
	expected	250.25	33.69	107.94	158.13

Notes: Count in each cell reports the number of the corresponding information sets that are observed. *trust*/*no trust*: subjects who invest in no less than/less than three rounds of Game 1. Person's chi-squared test: $p = 0.3204$; Two-sided Fisher's exact test: $p = 0.3308$; Spearman correlation is -0.048 with $p = 0.1779$.

Since the distribution of four categories of information sets that subjects observe in Game 2 is similar across the two groups of subjects, whether there is any difference in their strategies in Game 2 can be revealed through any difference between them in the percentage of *invest* action for each category of information set. Table 9 shows that the percentage of *invest* action in Game 2 is always higher for *trust* subjects than for *no trust* subjects in Game 1, which indicates that trustors' strategies in the two games are associated.

Table 9: Trustors' strategies across games

	H_1	H_2	H_3	H_4
trust subjects in Game 1	39.84%($\frac{49}{123}$)	25%($\frac{3}{12}$)	68.63%($\frac{35}{51}$)	98.44%($\frac{63}{64}$)
no trust subjects in Game 1	12.45%($\frac{30}{241}$)	5.41%($\frac{2}{37}$)	33.96%($\frac{36}{106}$)	95.18%($\frac{158}{166}$)
p-value	< 0.001	0.17	< 0.001	0.46

Notes: The percentage reports the fraction of *invest* action in Game 2. p-values are obtained from two-sided proportion test. p-values for H_2 and H_4 are calculated by simulation due to the small sample size.

Pattern 6 (Association of trustor's strategy in two games) *Trustor's strategy in Game 2 is associated with her strategy in Game 1. In particular, when observing information structure of low index the fraction of trustor subjects who invest is much higher for the group of trust subjects in Game 1 than for the group of no trust subjects in Game 1.*

6 Prosociality and strategic sophistication

Motivated by the observed treatment effects and six patterns, we propose in this section a model that is flexible to predict them. In particular, the proposed model enables to rationalize a designer's choosing the least informative structure. A main feature of the model is that we introduce heterogeneity in prosociality and strategic sophistication for trustee subjects, and introduce heterogeneity in prosociality, heterogeneity in viewpoint about opponents' prosociality and strategic sophistication for trustor subjects. This feature deviates from the standard assumption about selfish players and the notion of equilibrium.[32]

A1. Heterogeneity in prosociality

We assume that trustees are heterogeneous in the dimension of prosociality: a trustee may belong to selfish type or prosocial type. A trustee's payoff depends on the pecuniary payoff and his prosociality type $\theta_d \in \{\theta_{d1}, \theta_{d2}\}$: his payoff from the trustor's choosing *no invest* is 10, his payoff from the allocation according to plan $(22, 8)$ is 22, and his payoff from the allocation according to plan $(15, 15)$ is $15 + \theta_d$. For selfish trustee, $\theta_d = \theta_{d1} = 0$. For prosocial trustee, we assume that $\theta_d = \theta_{d2} > 7$, which naturally be interpreted as his gaining psychic payoff from making others better off. So a prosocial trustee prefers plan $(15, 15)$ to plan $(22, 8)$ while a selfish trustee prefers plan $(22, 8)$ to plan $(15, 15)$.

We assume that trustors are heterogeneous in the dimension of prosociality: a trustor may belong to selfish type or prosocial type. A trustor's payoff depends on the pecuniary payoff and her prosociality type $\theta_t \in \{\theta_{t1}, \theta_{t2}\}$: her payoff from choosing *no invest* is 10, and her payoff from choosing *invest* is the corresponding monetary payoff plus a value of θ_t, e.g., $15 + \theta_t$ or $8 + \theta_t$. For selfish trustor, $\theta_t = \theta_{t1} = 0$. For prosocial trustor, we assume that $\theta_t = \theta_{t2} > 2$, which naturally be interpreted as her gaining psychic payoff from placing trust on others. So a prosocial trustor prefers *invest* to *no invest* regardless of the trustee's choice.[33]

[32] An alternative modeling approach can be deviating from the assumption about selfishness but maintain the notion of equilibrium through introducing noises, e.g., quantal response equilibrium. Three considerations finally lead us to choose the present modeling approach over the alternative one. First, learning has been greatly suppressed in our experimental design and in turn subjects' actions are best interpreted as initial responses, which makes allowing for heterogeneous strategic sophistication a more plausible and realistic assumption. Second, in contrast with the noise interpretation about deviations from the standard equilibrium prediction, the structure of different levels of strategic sophistication provides more insights about our experimental data, which also coincide with observed behavioral patterns. Third, due to the existence of a continuum of equilbria a quantal response equilibrium analysis becomes rather technically involved and additionally one will have to rely on very restrictive assumptions about equilibrium selection when conducting structural estimation.

[33] Our specific assumption about payoff captures the notion of prosociality literally: a trustee in our game can determine only the pecuniary payoff allocation and a trustor can determine only

A2. Heterogeneity in viewpoint about opponents' prosociality

We assume that trustors are heterogeneous in their beliefs about trustees' prosociality, which may be interpreted that people hold different viewpoints on the fraction prosocial trustees in the population. In particular, a trustor with viewpoint π thinks that the fraction of prosocial trustees is π, or thinks that she is playing with a prosocial trustee with the chance of π and correspondingly playing with a selfish trustee with the remaining chance. We assume that π follows a uniform distribution over $[0, \overline{\pi}]$, where $\overline{\pi} \in (\frac{2}{7}, 1]$. For example, a trustor with viewpoint $\pi = 0$ thinks that she always faces a selfish trustee.

We could also assume that trustees are heterogeneous in their viewpoints about trustors' prosociality. Nevertheless, it can be shown that an introduction of heterogeneity in designers' viewpoint about trustors' prosociality at most enhances the flexibility of the model in a small sense. To economize the use of additional assumptions, we instead assume that all trustees hold the same viewpoint about trustors' prosociality, that is, think that the fraction of prosocial trustor is $\alpha \in (0, 1)$.

A3. Heterogeneity in strategic sophistication

We also assume that players are heterogeneous in the dimension of strategic sophistication by following a level k model structure: a L_k player believes that his/her opponent employs a L_{k-1} player's optimal strategy. Given our assumptions, a L_k trustee's belief about his opponent's strategy is specified as: a L_{k-1} prosocial trustor with viewpoint π follows her optimal strategy, a L_{k-1} selfish trustor with viewpoint π follows her optimal strategy, and he encounters with a prosocial trustor with the chance of α and the density of his encountering with a trustor with viewpoint $\pi \in [0, \overline{\pi}]$ is $\frac{1}{\overline{\pi}}$. For a L_k trustor with viewpoint π, her belief about

whether the (socially desirable) investment project is carried out, so a trustee's prosociality is best captured by assigning him an additional payoff once the allocation is beneficial to his opponent or equivalently he chooses to be trustworthy, and a trustor's prosociality is best captured by assigning her an additional payoff once she decides to conduct the (socially desirable) investment project or equivalently she chooses to place trust. In addition, the assumptions that $\theta_{d2} > 7$ and $\theta_{t2} > 2$ are made to generate a change in direction instead of a mere change in magnitude of prosocial players' behavior compared to selfish players.

her opponent's strategy includes: a L_{k-1} prosocial trustee follows his optimal strategy, a L_{k-1} selfish trustee follows his optimal strategy, and she encounters with a prosocial trustee with the chance of π.

We assume that a L_0 player makes decision as if he/she makes a non-strategic decision. Specifically, for L_0 players in Game 1 and also in Game 2, a prosocial trustee chooses $p = 1$, a selfish trustee chooses $p = 0$, a prosocial trustor chooses *invest* and a selfish trustor chooses *no invest*.[34] In Game 2, a L_0 trustee is also assumed to choose q_1 and q_2 independently according to the uniform distribution over $[0, 1]$. As in the literature, our assumption about L_0 player does not mean that there exists a L_0 player, and it serves to anchor initial belief of strategic consideration and it is mostly likely to exist only in the mind of L_1 players.

Trustor's prior and posterior

We now decompose a trustor's decision process into a few stages and formulate her prior and posterior about state 1 under these assumptions about heterogeneity. A trustor is assumed to form her prior about state 1 at a stage when the allocation plan has been determined according to p but information set has not be observed yet. A trustor is assumed to form her posterior about state 1 at a stage when information set has been observed (if any) and she is about to choose her action. Consider a trustor whose type is specified by her prosociality, her viewpoint about opponent's prosociality and her strategic sophistication level. Let E denote the event of state 1 or equivalently the determined allocation plan $(15, 15)$ and correspondingly E^c denote the event of state 2. Given (π, L_k), the trustor's prior

[34] Alternatively, we can assume that a L_0 selfish trustor's action is dependent on her viewpoint π, that is, choose *invest* for $\pi \in (\frac{2}{7}, \overline{\pi}]$ and choose *no invest* for $\pi \in [0, \frac{2}{7}]$. It is straightforward to show that the alternative assumption has no impact on L_1 trustee's optimal strategy.

about state 1 can be characterized as:

$$\begin{aligned}Pr(E) &= (1-\pi) * Pr(L_{k-1} \text{ selfish trustee's choice of } p) \\ &+ \pi * Pr(L_{k-1} \text{ prosocial trustee's choice of } p) \\ &\equiv (1-\pi) * p_{d1} + \pi * p_{d2}\end{aligned}$$

In Game 1 the trustor observes no information set and in turn her posterior is identical to her prior $Pr(E)$. In Game 2 the trustor's posterior after observing an information set $H = (q_1, q_2, s)$ is updated according to Bayes' law as follows:

$$\begin{aligned}Pr(E|H) &= Pr(E, d1|H) + Pr(E, d2|H) \\ &= \frac{Pr(E, d1, H) + Pr(E, d2, H)}{Pr(H)} \\ &= \frac{Pr(d1)Pr(E|d1)Pr(H|E, d1) + Pr(d2)Pr(E|d2)Pr(H|E, d2)}{Pr(H)} \\ &= \frac{(1-\pi)p_{d1}Pr(H|E, d1) + \pi p_{d2}Pr(H|E, d2)}{Pr(H)}\end{aligned}$$

where $Pr(H)$ can be explicitly written as,

$$\begin{aligned}Pr(H) &= Pr(E, d1, H) + Pr(E, d2, H) + Pr(E^c, d1, H) + Pr(E^c, d2, H) \\ &= (1-\pi)p_{d1}Pr(H|E, d1) + \pi p_{d2}Pr(H|E, d2) \\ &+ (1-\pi)(1-p_{d1})Pr(H|E^c, d1) + \pi(1-p_{d2})Pr(H|E^c, d2) \\ &= (1-\pi)p_{d1}Pr(q_1, q_2|d1)Pr(s|E, q_1, q_2) + \pi p_{d2}Pr(q_1, q_2|d2)Pr(s|E, q_1, q_2) \\ &+ (1-\pi)(1-p_{d1})Pr(q_1, q_2|d1)Pr(s|E^c, q_1, q_2) + \pi(1-p_{d2})Pr(q_1, q_2|d2)Pr(s|E^c, q_1, q_2).\end{aligned}$$

It is notable that when $Pr(H) = 0$ the above formula of specifying the trustor's posterior does not apply. The explicit expression of $Pr(H)$ hints that $Pr(H) = 0$ in the following situations for a trustor with (π, L_k): (1) observed information structure (q_1, q_2) is neither the optimal choice of L_{k-1} selfish trustee nor the optimal choice of L_{k-1} prosocial trustee, that is, $Pr(q_1, q_2|d1) = Pr(q_1, q_2|d2) = 0$; (2)

observed information structure (q_1, q_2) is consistent with the optimal choice of either L_{k-1} selfish trustee or L_{k-1} prosocial trustee but observed signal is inconsistent with the corresponding trustee's optimal choice of p, e.g., $Pr(q_1, q_2|d1) > 0, Pr(q_1, q_2|d2) = 0, Pr(s|E, q_1, q_2) = 0, p_{d1} = 1$. When a trustor observes a zero-probability information set, we have no a priori specification about her posterior.

A4. Conditionally pessimistic posterior in zero-probability information sets

On the one hand a trustor's posterior is ambiguous in zero-probability information sets. On the other hand, a complete specification of the trustor's optimal strategy requires to specify her optimal action in any information set, including the zero-probability information sets. This necessitates making a plausible assumption about a trustor's posterior in zero-probability information sets.[35] We assume that a trustor holds a conditionally pessimistic posterior in a zero-probability information set in the sense that she uses the "worst" posterior among all posteriors that are consistent with that information set. For instance, suppose $H = (q_1 = 0.8, q_2 = 0.4, s = b)$ is a zero-probability information set for a trustor. Her conditionally pessimistic posterior about state 1 is zero because any posterior from $[0, 1]$ is consistent with the information set and a posterior of zero is the "worst". Suppose $H = (q_1 = 0.8, q_2 = 0, s = b)$ is a zero-probability information set for a trustor. Her conditionally pessimistic posterior about state 1 is one because only a posterior of one is consistent with the information set. We argue that the assumption of conditionally pessimistic posterior in zero-probability information sets is a rather natural one because the trustor's optimal action based on such posterior is reminiscent of a maximin strategy, which is well motivated in game theory. In fact, a maxmin strategy dictates that players choose the strategy that

[35]Recall that in our previous equilibrium analysis we do not explicitly introduce the notion of prior and posterior. A belief of p there is conceptually equivalent to a prior about state 1 here. We apply the notion of subgame perfection to specify a belief of p on the off-equilibrium path, that is, applying the fixed point logic to specify a belief of p on those off-equilibrium (q_1, q_2). Nevertheless, the fixed point logic is dropped in the level k model so we have to take an alternative approach to specify a trustor's posterior in zero-probability information sets.

maximizes their expected payoffs in the "worst" one among all situations. Our assumption only makes a slight modification about the support of the "worst" situation: the support is restricted to all situations that are consistent with the observed information set.

Proposition 4. *Given the assumptions about players' heterogeneity in prosociality, viewpoint about prosociality, and strategic sophistication, and the assumption about conditionally pessimistic posterior in zero-probability information sets (A1-A4), the optimal strategy for each type is summarized in Table 10.*

TABLE 10: Optimal strategy for each behavioral type

Role	Type	Game 1	Game 2
trustee	(P, L_1)	$p = 1$	$p = 1, (q_1, q_2) \in [0,1] \times [0,1]$
	$(P, L_{k\geq 2})$	$p = 1$	$p = 1, (q_1, q_2) = (1,0)/(0,1)$
	(S, L_1)	$p = 0$	$p = 0, (q_1, q_2) \in [0,1] \times [0,1]$
	$(S, L_2/L_3)$	$p = 0$	$p = 0, q_1 = q_2$
	$(S, L_{k\geq 4})$	$p = 0$	$p = 1, (q_1, q_2) = (1,0)/(0,1)$
trustor	$(P, L_{k\geq 1}, \pi)$	$z = 1$	$z = 1$
	$(S, L_1/L_2, \pi)$	$z = 1$ if $\pi \in (\frac{2}{7}, \overline{\pi}]$; $z = 0$ if $\pi \in [0, \frac{2}{7}]$	$z = 1$ if $(\frac{q_1}{q_2} > \frac{2}{5}\frac{1-\pi}{\pi}, b)/(\frac{1-q_1}{1-q_2} > \frac{2}{5}\frac{1-\pi}{\pi}, w)$; $z = 0$ otherwise
	$(S, L_{k\geq 3}, \pi)$	same as above	$z = 1$ if $(0 < q_1 \leq 1, q_2 = 0, b)/(0 \leq q_1 < 1, q_2 = 1, w)$; $z = 0$ otherwise

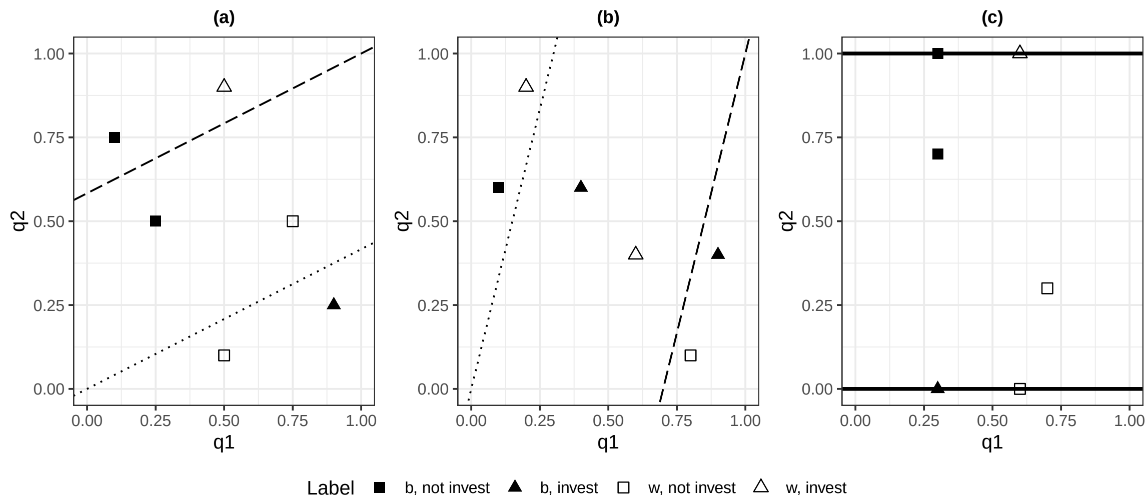
Notes: P and S index prosocial and selfish players respectively. Type L_0 specification: prosocial and selfish trustees choose $p = 1$ and $p = 0$ respectively, who also choose q_1 and q_2 independently in Game 2 according to the uniform distribution over $[0, 1]$; prosocial and selfish trustors choose *invest* and *not invest* respectively. When $\overline{\pi} < \frac{24(1-\alpha)}{49}$, $L_{k\geq 2}$ selfish trustees' strategy is $p = 1, (q_1, q_2) = (1, 0)/(0, 1)$.

Table 10 shows that subject types can only be separated coarsely based on optimal strategy. We illustrate the optimal strategy in Game 2 of trustors for three behaviorally separable type $(S, L_1/L_2, \pi < \frac{2}{7})$, $(S, L_1/L_2, \pi > \frac{2}{7})$ and $(S, L_{k\geq 3}, \pi)$ in Figure 4.

Remark 1. We discuss the role of the assumptions about heterogeneity in three dimensions. An introduction of prosocial trustee predicts a choice of $p = 1$ in both games. An introduction of prosocial trustor predicts an action of *invest* in both games.[36] An introduction of trustee's heterogeneity in strategic sophistication

[36]A prosocial trustor is not behaviorally separable from a selfish trustor with viewpoint $\pi \in (\frac{2}{7}, \overline{\pi}]$ in Game 1 but they are behaviorally separable in Game 2.

FIGURE 4: An illustration of trustor's optimal strategy



Notes: The dotted line is $\frac{q_1}{q_2} = \frac{2(1-\pi)}{5\pi}$ and the dashed line is $\frac{1-q_1}{1-q_2} = \frac{2(1-\pi)}{5\pi}$. (a): trustor of type (S, L_1/L_2,$\pi < \frac{2}{7}$); (b) trustor of type (S, L_1/L_2,$\pi > \frac{2}{7}$); (c): trustor of type (S, L_3,π).

enables to predict that: unsophisticated trustee subjects do not associate the choice of trustworthiness with the choice of information structure, moderately sophisticated subjects obfuscate their low trustworthiness through the least informative structure, and most sophisticated subjects expect that their opponents understand the message embedded in the least informative structure and consequently choose to be trustworthy with the most informative structure. An introduction of trustor's heterogeneity in strategic sophistication enables to predict that: unsophisticated trustor subjects make inference about the allocation plan through observed signal generated from observed information structure, and sophisticated subjects make inference about the allocation plan through both the message conveyed by information structure and observed signal generated from observed information structure.

Notably, the assumption that trustor's viewpoint about opponents' prosociality is heterogeneous and it takes value from a continuous support mainly serves to rationalize trustee's choice of the least informative structure, that is, explain why it is optimal for some trustee subjects to choose a strategy of $(p = 0, q_1 = q_2)$.

The general intuition is that when a trustee thinks that his opponent is not aware of the association between trustworthiness and information structure he has an incentive to obfuscate his choice of $p = 0$ through using an information structure of $(q_1 = q_2)$. Specifically, consider trustors who fail to realize that information structure per se conveys a message about trustworthiness. We know that a trustor with a posterior exceeding a cutoff value chooses *invest* and that her posterior about state 1 is increasing in their prior, which itself is increasing in their viewpoint π. On the other hand, for an information set (q_1, q_2, b) a trustor's posterior is also increasing in $\frac{q_1}{q_2}$. Thus, as $\frac{q_1}{q_2}$ increases those trustors with a lower viewpoint π start to choose *invest*, i.e., lower the cutoff value of π, which implies that a higher $\frac{q_1}{q_2}$ increases the chance of *invest* because the assumption about viewpoint π guarantees that the fraction of trustors with viewpoint exceeding a cutoff value is decreasing in the cutoff value. For a similar reason, a higher $\frac{1-q_1}{1-q_2}$ increases the chance of *invest* for an information set (q_1, q_2, w). In turn, for a trustee who chooses $p = 0$, he faces a trade-off between choosing a high $\frac{q_1}{q_2}$ and choosing a high chance of generating a black signal (equivalently choosing a high q_2 given $p = 0$), or he faces a trade-off between choosing a high $\frac{1-q_1}{1-q_2}$ and choosing a high chance of generating a white signal (equivalently choosing a high $1 - q_2$ given $p = 0$). Thus, it is clear now the role of the assumption about π is to make the chance of a trustor's choosing *invest* (or equivalently the fraction of trustors choosing *invest*) be responsive to the change of information structure and consequently create a trade-off problem for the trustee who chooses $p = 0$.

Remark 2. We finally discuss how the behavioral model predicts the treatment effects and behavioral patterns reported in Section 5. As Table 10 demonstrates, a trustee is likely to choose $p = 1$ in both games, choose $p = 0$ in both games or change from $p = 0$ in Game 1 to $p = 1$ in Game 2 depending on his type, which is consistent with Pattern 2 and in turn the treatment effect about trustworthiness. In Game 2, a trustee is likely to choose $(p = 0, q_1 = q_2)$ or

$(p=1,q_1=1,q_2=0)/(p=1,q_1=0,q_2=1)$, which is consistent with Pattern 1 and Pattern 3. In Game 2, a trustor is likely to make inference about the underlying state from observed signal generated from observed information structure, e.g., type $(S, L_1/L_2, \pi)$, or make inference by further incorporating the message from information structure, e.g., type $(S, L_{k\geq 3}, \pi)$, which is consistent with Pattern 4 and Pattern 5. In addition, the model predicts that whether an introduction of information design boosts trust depends on the fraction of trustors with viewpoint exceeding $\frac{2}{7}$ and trustees' choice of information structure and trustworthiness: if the fraction is small and trustees choose high likelihood ratio $\frac{q_1}{q_2}/\frac{1-q_1}{1-q_2}$ with high trustworthiness, it indeed increases trust; otherwise an introduction of information design is likely to backfire.[37]

7 Type estimation and orthogonality

In this section, we employ the maximum likelihood estimation method to estimate each subject's type based on the proposed behavioral model. We also look at the degree of consistency between a subject's econometric type and two other classifications: type exactly prescribed by the theory and type based on an intuitive criterion. Since a subject's type consists of three dimensions: prosociality, strategic sophistication and his/her viewpoint about opponents' prosociality, we additionally explore statistical independence between the three dimensions and find that a subject's prosociality is independent of his/her strategic sophistication level.

[37]The sign of change in the probability of trusting act from Game 1 to Game 2 varies across trustor types. $(S, L_1/L_2, \pi \leq \frac{2}{7}) : \Delta = Pr(\{\frac{q_1}{q_2} > \frac{2}{5}\frac{1-\pi}{\pi}, b\} \cup \{\frac{1-q_1}{1-q_2} > \frac{2}{5}\frac{1-\pi}{\pi}, w)\} > 0$; $(S, L_1/L_2, \pi > \frac{2}{7}) : \Delta = Pr(\{\frac{q_1}{q_2} > \frac{2}{5}\frac{1-\pi}{\pi}, b\} \cup \{\frac{1-q_1}{1-q_2} > \frac{2}{5}\frac{1-\pi}{\pi}, w)\} - 1 < 0$; $(S, L_3, \pi \leq \frac{2}{7}) : \Delta = Pr(\{0 < q_1 \leq 1, q_2 = 0, b\} \cup \{0 \leq q_1 < 1, q_2 = 1, w)\} > 0$;$(S, L_3, \pi > \frac{2}{7}) : \Delta = Pr(\{0 < q_1 \leq 1, q_2 = 0, b\} \cup \{0 \leq q_1 < 1, q_2 = 1, w)\} - 1 < 0$.

7.1 Econometric type estimation

Let y index a generic combination of values in the dimensions of prosociality and strategic sophistication. Since trustees are assumed to hold the same viewpoint about opponents' prosociality, characterized by α, a trustee's type can be captured by y. A trustor's type is determined by (y, π). The predicted action $c_{(y,g)}$ of a trustee with type y specifies his optimal action in Game $g(g \in \{1, 2\})$, where $c_{(y,g)} = p \in [0, 1]$ in Game 1 and $c_{(y,g)} = (p, q_1, q_2) \in [0, 1] \times [0, 1] \times [0, 1]$ in Game 2. The predicted action $c_{(y,\pi,g)}(H)$ of a trustor with type (y, π) specifies her optimal action in Game g when facing an information set H, where $H \in \emptyset$ in Game 1, $H = (q_1, q_2, s) \in [0, 1] \times [0, 1] \times \{b, w\}$ in Game 2, and $c_{(y,\pi,g)}(H) = z \in \{1, 0\}$.

Let Ω index the action space. For trustees, $\Omega = \{p : p \in [0, 1]\}$ in Game 1 and $\Omega = \{(p, q_1, q_2) : p \in [0, 1], q_1 \in [0, 1], q_2 \in [0, 1]\}$ in Game 2. For trustors, $\Omega = \{1, 0\}$ in both games. Let $V_{(y,g)}(c_g)$ index a trustee's expected payoff from choosing action c_g in Game g and $V_{(y,\pi,g)}(c_g|H)$ index a trustor's expected payoff from choosing action c_g when observing an information set H in Game g. Table 11 presents the expected payoff functions of different player types. Compared to types that are separated by differences in optimal strategies as in Table 10, more types are further differentiated by differences in expected payoff functions since those types can be weakly separated by differences in the deviation costs implied by their beliefs although they cannot be separated based on the optimal strategy. As in the literature, L_0 type is assumed to only exist in the mind of L_1 type and we will not put L_0 type in the list of candidate types when conducting structural estimation.

We assume that each subject of a certain type normally follows the predicted action of the type but subject to the standard logit error structure. Let λ index the precision parameter of the standard logit error structure. A trustee with type y and a trustor with type (y, π) observing H choose an action c_g with the probability

Table 11: Expected payoff function of different types

Role	Type	Game 1	Game 2
trustee	(P, L_1)	$p*\{(15+\theta_{d2})\alpha+(1-\alpha)[P_r(\pi>\frac{2}{7})(15+\theta_{d2})+10P_r(\pi<\frac{2}{7})]\}$ $+(1-p)*\{22\alpha+(1-\alpha)[22P_r(\pi>\frac{2}{7})+10P_r(\pi<\frac{2}{7})]\}$	same as Game 1
	$(P, L_2/L_3)$	same as (P, L_1)	$p*\{10+(5+\theta_{d2})\alpha+(5+\theta_{d2})(1-\alpha)[q_1P_r(\pi>\frac{0.4}{0.4+\frac{q_1}{q_2}})+(1-q_1)P_r(\pi>\frac{0.4}{0.4+\frac{1-q_1}{1-q_2}})]\}$ $+(1-p)*\{10+12\alpha+12(1-\alpha)[q_2P_r(\pi>\frac{0.4}{0.4+\frac{q_1}{q_2}})+(1-q_2)P_r(\pi>\frac{0.4}{0.4+\frac{1-q_1}{1-q_2}})]\}$
	$(P, L_{k\geq4})$	same as (P, L_1)	$p*\{(15+\theta_{d2})\alpha+(1-\alpha)[q_1(15+\theta_{d2})+10(1-q_1)]\}+(1-p)*(12\alpha+10)$ if $(q_1,q_2)\in\phi_1$; $p*\{(15+\theta_{d2})\alpha+(1-\alpha)[(1-q_1)(15+\theta_{d2})+10q_1]\}+(1-p)(12\alpha+10)$ if $(q_1,q_2)\in\phi_2$; $p[(15+\theta_{d2})\alpha+10(1-\alpha)]+(1-p)(12\alpha+10)$ otherwise
	(S, L_1)	$p*\{15\alpha+(1-\alpha)[15P_r(\pi>\frac{2}{7})+10P_r(\pi<\frac{2}{7})]\}$ $+(1-p)*\{22\alpha+(1-\alpha)[22P_r(\pi>\frac{2}{7})+10P_r(\pi<\frac{2}{7})]\}$	same as Game 1
	$(S, L_2/L_3)$	same as (S, L_1)	$p*\{10+5\alpha+5(1-\alpha)[q_1P_r(\pi>\frac{0.4}{0.4+\frac{q_1}{q_2}})+(1-q_1)P_r(\pi>\frac{0.4}{0.4+\frac{1-q_1}{1-q_2}})]\}$ $+(1-p)*\{10+12\alpha+12(1-\alpha)[q_2P_r(\pi>\frac{0.4}{0.4+\frac{q_1}{q_2}})+(1-q_2)P_r(\pi>\frac{0.4}{0.4+\frac{1-q_1}{1-q_2}})]\}$
	$(S, L_{k\geq4})$	same as (S, L_1)	$p*\{15\alpha+(1-\alpha)[15q_1+10(1-q_1)]\}+(1-p)(12\alpha+10)$ if $(q_1,q_2)\in\phi_1$; $p*\{(15\alpha+(1-\alpha)[15(1-q_1)+10q_1]\}+(1-p)(12\alpha+10)$ if $(q_1,q_2)\in\phi_2$; $p(10+5\alpha)+(1-p)(12\alpha+10)$ otherwise
Trustor	$(P, L_1/L_2, \pi)$	$(15+\theta_{t2})\pi+(8+\theta_{t2})(1-\pi)$ if $z=1$; 10 if $z=0$	$8+\theta_{t2}+7\frac{\pi q_1}{\pi q_1+(1-\pi)q_2}$ if $(s=b, z=1)$; $8+\theta_{t2}+7\frac{\pi(1-q_1)}{\pi(1-q_1)+(1-\pi)(1-q_2)}$ if $(s=w, z=1)$ 10 if $z=0$
	$(P, L_{k\geq3}, \pi)$	$(15+\theta_{t2})\pi+(8+\theta_{t2})(1-\pi)$ if $z=1$; 10 if $z=0$	$15+\theta_{t2}$ if $(q_1,q_2,s)\in\hat{H}$ &$z=1$; $8+\theta_{t2}$ if $(q_1,q_2,s)\notin\hat{H}$&$z=1$ 10 if $z=0$
	$(S, L_1/L_2, \pi)$	$15\pi+8(1-\pi)$ if z=1; 10 if z=0	$8+7\frac{\pi q_1}{\pi q_1+(1-\pi)q_2}$ if $(s=b, z=1)$; $8+7\frac{\pi(1-q_1)}{\pi(1-q_1)+(1-\pi)(1-q_2)}$ if $(s=w, z=1)$ 10 if $z=0$
	$(S, L_{k\geq3}, \pi)$	$15\pi+8(1-\pi)$ if z=1; 10 if z=0	15 if $(q_1,q_2,s)\in\hat{H}$ &$z=1$; 8 if $(q_1,q_2,s)\notin\hat{H}$&$z=1$ 10 if $z=0$

Notes: $\phi_1 \equiv (0 < q_1 \leqslant 1, q_2 = 0)$; $\phi_2 \equiv (0 \leq q_1 < 1, q_2 = 1)$; $\hat{H} \equiv \{(\phi_1, s = b)\} \cup \{(\phi_2, s = w)\}$.

specified as follows.

$$P^d_{(y,g)}(c_g) = \frac{exp[\lambda V_{(y,g)}(c_g)]}{\int_\Omega exp[\lambda V_{(y,g)}(\tilde{c})]\, d\tilde{c}}\text{ , for trustee} \tag{1}$$

$$P^t_{(y,\pi,g)}(c_g|H) = \frac{exp[\lambda V_{(y,\pi,g)}(c_g|H)]}{\sum_\Omega exp[\lambda V_{(y,\pi,g)}(\tilde{c}|H)]}\text{ , for trustor} \tag{2}$$

As usual, the choice probability is increasing in the expected payoff of the corresponding action and the dispersion of choice declines with the strength of payoff incentives. In addition, subject's choice approaches uniform randomness as $\lambda \to 0$ and approaches the predicted action as $\lambda \to \infty$.

An observation sample for a subject includes his/her observed information sets (if any) and actions in ten experimental rounds (five rounds $\times$ two games). Let (c_i, H_i) index subject i's sample, where $c^i = \{c^i_{gr}\}_{g\in\{1,2\},r\in\{1,\cdots,5\}}$ and $H^i = \{H^i_{gr}\}_{g\in\{1,2\},r\in\{1,\cdots,5\}}$. With the assumption that a subject's making errors are independent across ten rounds, the likelihood of observing a subject's sample

(c^i, H^i) conditional on the subject's type can be specified as:

$$L^d_y(c^i) = \prod_{g=1}^{2}\prod_{r=1}^{5} P^d_{(y,g)}(c^i_{gr}) = \prod_{g=1}^{2}\prod_{r=1}^{5} \frac{exp[\lambda V_{(y,g)}(c^i_{gr})]}{\int_\Omega exp[\lambda V_{(y,g)}(\tilde{c})]\, d\tilde{c}} \tag{3}$$

$$L^t_{(y,\pi)}(c^i, H^i) = \prod_{g=1}^{2}\prod_{r=1}^{5} P^t_{(y,\pi,g)}(c^i_{gr}) = \prod_{g=1}^{2}\prod_{r=1}^{5} \frac{exp[\lambda V_{(y,\pi,g)}(c^i_{gr}|H)]}{\sum_\Omega exp[\lambda V_{(y,\pi,g)}(\tilde{c}|H)]} \tag{4}$$

Let β_y be the prior probability of a trustee being type y and $\beta_{(y,\pi)}$ be the prior density of a trustor being type (y, π). According to Table 11, a trustee's type may belong to one of six categories and for a trustor with viewpoint π her type may belong to one of four catogories. Thus, we have $\sum_{y=1}^{6} \beta_y = 1$ and $\sum_{y=1}^{4} \int_0^{\bar{\pi}} \beta_{(y,\pi)}\, d\pi = 1$. Then the likelihood of observing a subject's sample (c^i, H^i) unconditional on type can be formulated as:

$$L^d(c^i) = \sum_{y=1}^{6} \beta_y \prod_{g=1}^{2}\prod_{r=1}^{5} \frac{exp[\lambda V_{(y,g)}(c^i_{gr})]}{\int_\Omega exp[\lambda V_{(y,g)}(\tilde{c})]\, d\tilde{c}}] \tag{5}$$

$$L^t(c^i, H^i) = \sum_{y=1}^{4} \int_0^{\bar{\pi}} \beta_{(y,\pi)} \prod_{g=1}^{2}\prod_{r=1}^{5} \frac{exp[\lambda V_{(y,\pi,g)}(c^i_{gr}|H)]}{\sum_\Omega exp[\lambda V_{(y,\pi,g)}(\tilde{c}|H)]}\, d\pi] \tag{6}$$

Since we perform analysis at the individual level, the precision parameter λ is implicitly assumed to be subject-specific. In addition, since the two games are rather different we assume that λ is game-specific: $\{\lambda_g\}_{g=1}^2$. In the subject-by-subject analysis, we jointly estimate a subject's game-specific λ_g and type probabilities. Since the likelihood function $L^d(c^i)$ is linear in β_y, the maximum likelihood estimate of a trustee's type probabilities sets $\beta_y = 1$ for the (generically unique) y that yields the highest $L^d_y(c^i)$, which is obtained by maximizing $L^d_y(c^i)$ over $\{\lambda_g\}_{g=1}^2$ given type y. For a similar reason, the maximum likelihood estimate of a trustor's type density sets $\beta_{(y,\pi)} = 1$ for the (generically unique) (y, π) that yields the highest $L^t_{(y,\pi)}(c^i, , H^i)$, which is obtained by maximizing $L^t_{(y,\pi)}(c^i, H^i)$ over $\{\lambda_g\}_{g=1}^2$ given type (y, π).

It is a prerequisite to specify the values of background parameters $(\theta_{d2}, \theta_{t2}, \alpha, \bar{\pi})$ before we can conduct maximum likelihood estimation of each subject's type and game-specific λ_g. There is no a priori requirement on the pick of the values of these parameters. We assume that the background parameters are role specific and apply the spirit of the maximum likelihood estimation to pick their values. Specifically, for each value of $(\theta_{d2}, \theta_{t2}, \alpha, \bar{\pi})$ taken from a four-dimensional range we conduct the maximum likelihood estimation of each subject's type and precision parameter. We sum up the obtained likelihood values over trustee subjects and over trustor subjects separately, and then the value of $(\theta_{d2}, \theta_{t2}, \alpha, \bar{\pi})$ that maximizes the corresponding sum is picked as the corresponding value of background parameters.

In practice, our estimation procedure proceeds as follows. We first specify a discretized range of background parameters, e.g., $(\theta_{d2}, \theta_{t2}, \alpha, \bar{\pi}) \in \{7.5, 8\} \times \{2.5, 3\} \times \{0, 0.2, 0.4, 0.6, 0.8, 1\} \times \{0.15, 0.3, 0.45, 0.6, 0.75, 0.9\}$. Then for each $(\theta_{d2}, \theta_{t2}, \alpha, \bar{\pi})$ taken from this range, we conduct subject by subject analysis and maximize $lnL_y^d(c^i)$ and $lnL_y^t(c^i)$ over type and precision parameter. The value of $(\theta_{d2}, \theta_{t2}, \alpha, \bar{\pi})$ is picked for trustees and trustors respectively in a maximum likelihood fashion. Finally, each subject's type is estimated given the picked value of $(\theta_{d2}, \theta_{t2}, \alpha, \bar{\pi})$.[38]

We report each subject's type estimate, the corresponding precision parameter estimates, and standard errors of the parameters in Appendix B.[39] We also perform likelihood ratio tests at the subject level on the null hypotheses of $\lambda_1 = \lambda_2 = 0$ and $\lambda_1 = \lambda_2$, which approximate a random choice model and a constant precision

[38] As a robust check, we also vary the values of background parameters from the picked one and find that proportion of types remain more or less the same, which indicates that a subject's estimated type roughly remain unchanged when the values of background parameters deviate from the picked one. Details about the robust check are reported in Appendix B.

[39] Standard errors of the parameters in our model are obtained using a jackknife procedure, which can be roughly described as a "leave-one-out" procedure. In each run of jackknife, we re-estimate each subject's parameters with one of his/her observations (10 in total for any subject) excluded, while holding the subject's estimated type that is based on all observations. Then the standard deviations of the parameter estimates across all the runs are the standard errors of the parameters.This jackknife procedure has also been used in Crawford and Iriberri (2007b).

parameter between the two games respectively. For trustee subjects, the null hypothesis of $\lambda_1 = \lambda_2 = 0$ is rejected for 89.4% (84.4%) of them at the significance level of five percent (one percent), which suggests that for most of trustee subjects the proposed model is favored against a random choice model. The null hypothesis of $\lambda_1 = \lambda_2$ is rejected for 44.4% (32.5%) of them at the significance level of five percent (one percent), which suggests that the assumption about game-specific precision parameter is necessary for a considerable amount of trustee subjects. For trustor subjects, the null hypothesis of $\lambda_1 = \lambda_2 = 0$ is rejected for 79.8% (50.8%) of them at the significance level of five percent (one percent), which suggest that for the majority of trustor subjects the proposed model is favored against a random choice model. The null hypothesis of $\lambda_1 = \lambda_2$ is rejected for 5% (1.9%) of them at the significance level of five percent (one percent), which suggests that the assumption about game-specific precision parameter seems to be not necessary for trustor subjects.

7.2 Comparison of econometric type and intuitive type

We now investigate to what extent a subject's type based on econometric estimation coincides with his/her type that is based on an exact criterion from the model and the type that is based on an intuitive criterion.

We define the exact criterion of classifying a subject's type as a requirement that the subject's action in each of ten experimental rounds follows exactly the theoretically predicted action. Based on the exact criterion, a trustee subject is classified as one of six specific types or is unclassified; and a trustor subject is classified as one of eight specific types or is unclassified. A trustee is always classified as a unique type based on the exact criterion. However, a trustor is likely to be classified as more than one types based on the exact criterion for two reasons below. First, prosocial trustor subjects with different strategic sophostication levels and with different viewpoints about others' prosociality cannot be distinguished behaviorally per se. Second, while type $(S, L_1/L_2, \pi \leq \frac{2}{7})$ and trustor type $(S, L_{k\geq 3}, \pi \leq \frac{2}{7})$ can be behaviorally separated when the sample size is large enough, we may not be able to separate them when the sample size is small as in the case of our having only five round observations of Game 2. For instance, if the five observed information sets for a subject are all $(0 < q_1 \leq 1, q_2 = 0, b)/(0 \leq q_1 < 1, q_2 = 1, w)$ and she always chooses *invest* in these five rounds, then she can be classified as both type $(S, L_1/L_2, \pi \leq \frac{2}{7})$ and trustor type $(S, L_{k\geq 3}, \pi \leq \frac{2}{7})$.

We define the intuitive criterion of classifying a subject's type as a requirement that lifts the very stringent condition of the exact criterion and meantime maintains some of its intuitive features. Specifically, for trustee subjects the conditions mainly include: whether the mean value of p in five rounds of a game exceeds 0.5 or does not exceeds 0.5, and whether the mean value of Blackwell's informativeness in five rounds of Game 2 exceeds 0.7 or falls below 0.3. Since the predicted action for trustor subjects is binary, we cannot apply a similar approach to lift the condition of the exact criterion for trustor subjects. Instead, we apply an alternative idea of rationalization: identify the minimum number of observations that must be removed from a subject to guarantee that all the remaining observations from the subject are consistent with the theoretically predicted action. The specific procedure proceeds as follows. We identify the number of observations that must be removed over each type we hypothetically classify the subject to. If at most three observations must be removed for a certain type and the number of observation removal is the smallest compared to for any other type, then we classify the subject as this specific type based on the intuitive criterion; otherwise we classify the subject as unclassified. We summarize the two criteria of type classification in Table 12.

Table 13 summarizes subjects' type distributions based on five criteria: exact criterion, intuitive criterion, econometric criterion, and econometric criterion ex-

Table 12: Type classification based on exact Criterion and intuitive criterion

Role	Type	Exact Criterion	Intuitive Criterion
trustee	(P, L_1)	$p^1 = p^2 = 1, \sigma_{\lvert q_1 - q_2 \rvert} > 0$	$\overline{p^1} > 0.5, \overline{p^2} > 0.5, \sigma_{\lvert q_1 - q_2 \rvert} > 0$
	$(P, L_2/L_3)$ $(P, L_{k \geq 4})$	$p^1 = p^2 = 1, \lvert q_1 - q_2 \rvert = 1$	$\overline{p^1} > 0.5, \overline{p^2} > 0.5, \overline{\lvert q_1 - q_2 \rvert} \geq 0.7$
	(S, L_1)	$p^1 = p^2 = 0, \sigma_{\lvert q_1 - q_2 \rvert} > 0$	$\overline{p^1} \leq 0.5, \overline{p^2} \leq 0.5, \sigma_{\lvert q_1 - q_2 \rvert} > 0$
	$(S, L_2/L_3)$	$p^1 = p^2 = 0, \lvert q_1 - q_2 \rvert = 0$	$\overline{p^1} \leq 0.5, \overline{p^2} \leq 0.5, \overline{\lvert q_1 - q_2 \rvert} \leq 0.3$
	$(S, L_{k \geq 4})$	$p^1 = 0, p^2 = 1, \lvert q_1 - q_2 \rvert = 1$	$\overline{p^1} \leq 0.5, \overline{p^2} > 0.5, \overline{\lvert q_1 - q_2 \rvert} \geq 0.7$
	Unclassified	otherwise	otherwise
trustor	$(P, L_1/L_2, \pi \leq \frac{2}{7})$ $(P, L_1/L_2, \pi > \frac{2}{7})$ $(P, L_{k \geq 3}, \pi \leq \frac{2}{7})$ $(P, L_{k \geq 3}, \pi > \frac{2}{7})$	$z^1 = z^2 = 1$	Rationalizable if at most 3 obs removed
	$(S, L_1/L_2, \pi \leq \frac{2}{7})$	$z^1 = 0, z^2 = 1$ iff $(\frac{q_1}{q_2} > \frac{0.4(1-\pi)}{\pi}, b)/(\frac{1-q_1}{1-q_2} > \frac{0.4(1-\pi)}{\pi}, w)$	
	$(S, L_1/L_2, \pi > \frac{2}{7})$	$z^1 = 1, z^2 = 1$ iff $(\frac{q_1}{q_2} > \frac{0.4(1-\pi)}{\pi}, b)/(\frac{1-q_1}{1-q_2} > \frac{0.4(1-\pi)}{\pi}, w)$	
	$(S, L_{k \geq 3}, \pi \leq \frac{2}{7})$	$z^1 = 0, z^2 = 1$ iff $(0 < q_1 \leq 1, q_2 = 0, b)/(0 \leq q_1 < 1, q_2 = 1, w)$	
	$(S, L_{k \geq 3}, \pi > \frac{2}{7})$	$z^1 = 1, z^2 = 1$ iff $(0 < q_1 \leq 1, q_2 = 0, b)/(0 \leq q_1 < 1, q_2 = 1, w)$	
	Unclassified	otherwise	otherwise

Note: $p^i = j, z^i = j$ $(i = 1/2, j = 0/1)$ refers to the values of p and z in each of five rounds in game i. The value with a bar refers to the mean value in five founds of a game. $\lvert q_1 - q_2 \rvert$ refers to the index of Blackwell's informativeness in each of five rounds in game 2 and $\sigma_{\lvert q_1 - q_2 \rvert}$ refers to the corresponding standard deviation across the five rounds. The requirement *Rationalized with at most three observations removed* includes two parts: (1) the remaining observations can be predicted by the corresponding exact type after we remove at most three observations; and (2) the number of removed observations is the smallest when being classified as this type compared to when being classified as any other type.

cluding those which we cannot reject a random choice model at the significance levels of five percent and one percent. We find that econometric type (type based on econometric estimation) is largely consistent with exact type (type based on the exact criterion) and intuitive type (type based on the intuitive criterion). Specifically, for those trustee subjects who are assigned exact type, the estimated type based on three econometric criteria is consistent with the exact type 86.7% ($\frac{26}{30}$), 86.7% and 86.7% of the time respectively. [40] For those trustor subjects who are assigned exact type(s), the estimated type based on three econometric criteria is consistent with the exact type 100% ($\frac{57}{57}$), 100% and 100% of time respectively.

[40] Four subjects, labeld as 206, 304, 1318 and 1319, are estimated as type $(S, L_2/L_3)$ but classified as type (S, L_1) based on the exact criterion. Subject 206 has $\lvert q_1 - q_2 \rvert = 0$ in 4 but not 5 rounds; and subjects 304, 1318 and 1319 have $\lvert q_1 - q_2 \rvert \leqslant 0.05$ in at least 4 rounds. Recall that type $(S, L_2/L_3)$ based on the exact criterion requires $\lvert q_1 - q_2 \rvert = 0$ in each round, which is a very stringent condition. In addition, the unit of increment of p, q_1 and q_2 we picked is 0.05 in our econometric estimation. So the inconsistency for these 4 subjects seems to be acceptable.

For those designer subjects who are assigned intuitive type(s), the estimated type based on three econometric criteria is consistent with their intuitive type 79.4% ($\frac{108}{136}$), 81.1% ($\frac{103}{127}$) and 82.6% ($\frac{100}{121}$) of time respectively. For those trustor subjects who are assigned intuitive type(s), the estimated type based on three econometric criteria is consistent with the intuitive type 92.1% ($\frac{140}{152}$), 97.5% ($\frac{116}{119}$) and 100% ($\frac{65}{65}$) of time respectively. The high consistency shows that econometric estimation indeed confirms intuitive analysis of subjects' types and additionally refine our analysis about subjects' types when intuition does not provide a clear guidance.

TABLE 13: Summary of subjects' estimated type distributions

Role	Type	Exact	Intuitive	Econometric	Econometric, excluding random
trustee	(P, L_1)	0	16	10	4 (1)
	$(P, L_2/L_3)$	2	8	13	10 (8)
	$(P, L_{k\geq 4})$			9	7 (7)
	(S, L_1)	12	77	27	25 (25)
	$(S, L_2/L_3)$	3	65	52	49 (47)
	$(S, L_{k\geq 4})$	13	27	49	48 (47)
	Unclassified	130	24	0	17 (25)
trustor	$(P, L_1/L_2, \pi \leq \frac{2}{7})$	4	26	12	9 (4)
	$(P, L_1/L_2, \pi > \frac{2}{7})$			0	0 (0)
	$(P, L_{k\geq 3}, \pi \leq \frac{2}{7})$			3	0 (0)
	$(P, L_{k\geq 3}, \pi > \frac{2}{7})$			1	0 (0)
	$(S, L_1/L_2, \pi \leq \frac{2}{7})$	47	115	106	91 (53)
	$(S, L_1/L_2, \pi > \frac{2}{7})$	7	30	17	8 (4)
	$(S, L_{k\geq 3}, \pi \leq \frac{2}{7})$	33	103	42	37 (33)
	$(S, L_{k\geq 3}, \pi > \frac{2}{7})$	3	34	12	9 (4)
	Unclassified	103	8	0	39 (95)

Notes: Columns 3-6 report respectively the number of each subject type when the type is based on an exact criterion, an intuitive criterion, an econometric estimation, and an econometric estimation with additionally the null hypothesis that $\lambda_1 = \lambda_2 = 0$ being rejected at the significance level of five percent (one percent). In each column, the total number of trustee (trustor) subjects over types may exceed the number of trustee (trustor) subjects because a subject may be classified as more than one type except for classifying a trustee subject based on the exact criterion.

7.3 Orthogonality of prosociality and strategic sophistication

Since a subject's type is characterized by two dimensions (for trustee) or by three dimensions (for trustor), we finally explore statistical independence between the two or three dimensions, e.g., whether a subject's prosociality is independent of the subject's strategic sophistication level. Based on econometric type estimation, we report the association between a trustee subject's prosociality and strategical sophistication in Table 14 and the association between a trustor subject's prosociality and strategical sophistication in Table 15. The main message is that a subject's prosociality is not associated with his/her strategical sophistication level. Additionally, we find that a trustor subject's viewpoint about designers' prosociality is not associated with her prosociality and it is postively associated with her strategic sophistication level, which is reported in Appendix B.[41]

Table 14: Association between Prosociality and Strategic Sophistication: Trustee

Prosociality		Strategic sophistication L_1	L_2/L_3	$L_{k\geq4}$	Total
Prosocial	count	10	13	9	32
	expected	7.4	13	11.6	
Selfish	count	27	52	49	128
	expected	29.6	52	46.4	
Total		37	65	58	160

Notes: The two-way table presents the number of subjects for each of six trustee types. Pearson's Chi-squared test: $p = 0.3925$; Two-sided Fisher's exact test: $p = 0.3802$; and Spearman correlation is -0.1059 with $p = 0.1826$.

[41] When we restrict the analysis to the subsample of subjects who are classified as an econometric type excluding those which we cannot reject a random choice model at the significance level of five percent, all the messages about statistical independence between dimensions remain the same.

Table 15: Association between Prosociality and Strategic Sophistication: Trustor

Prosociality		Strategic sophistication L_1/L_2	$L_{k\geq 3}$	Total
Prosocial	count	12	4	16
	expected	12.85	3.15	
Selfish	count	90	21	111
	expected	89.15	21.85	
Total		102	25	127

Notes: The two-way table presents the number of trustor subjects based on the classification of prosociality and strategical sophistication. Pearson's Chi-squared test: $p = 0.8137$; Two-sided Fisher's exact test: $p = 0.5182$; and Spearman correlation is 0.0507 with $p = 0.571$. 33 out of 160 trustor subjects are excluded from this analysis since they are classified as more than one types.

8 Conclusion

This article explores a realistic market innovation to foster trusting act and addresses the question of whether trustees' taking the lead and designing information about trustworthiness indeed boosts trusting act. We design two experimental games: in the control condition trustees moves first without any tool of signaling their trustworthiness while in the treatment condition trustees moves first and have access to signaling their trustworthiness by designing an information structure. Experimental evidence shows that an introduction of such information design substantially enhances both trustworthiness and trusting act.

This article also investigates the mechanism underlying the treatment effect, that is, explains why the market innovation of information design foster trusting act. We first propose an equilibrium model under the standard assumptions. When having no access to any tool, trustors' foreseeing their being exploited by trustees drives out any trusting act. When having access to a tool of information design, trustees' internalizing trustors' optimal responses to their information design finally drives both parties to a fixed point where trustors place conditional trusting act and trustees choose to be trustworthy with a positive (up to full) probability. Since the equilibrium model only fits the treatment effects qualitatively but not quantitatively and additionally it fails to capture several observed behavioral patterns, we postulate an alternative model accounting for experimental findings that allows for heterogeneity in the dimensions of prosociality, viewpoint about opponents' prosociality and strategic sophistication. The alternative model enables to generate the treatment effects and observed patterns, and particularly it rationalizes some trustees's choosing to be not trustworthy with the least informative structure as an optimal response to some less sophisticated trustors who fail to understand the message conveyed by the information structure per se.

In addition to its capability of generating observed patterns, the alternative model also receives support econometrically. We structurally estimate each player subject's behavioral type based on the behavioral model and find that for most of subjects the proposed model is favored against a random choice model. In addition, we find that subject's type estimation almost fully coincides with the type based on the exact prescription of the model and is largely consistent with the type based on an intuitive classification criterion.

As a byproduct, this article also provides evidence that the dimension of prosociality is statistically independent of the dimension of strategic sophistication. This finding confirms that the implicitly independent assumption about the two dimensions in the proposed behavioral model is plausible. Since this article is not designed to explore the relationship between them, we view the evidence to be suggestive rather than conclusive. An investigation of the relationship is of independent interest to social contexts in which both prosociality and strategic sophistication are likely to play a role and we leave it for future work.

Overall, among the listed contributions to the four strands of literature in Section 2, we view this article primarily contributes to the literature concerning trust. Resolving the social dilemma involving trust is a broad question that can be addressed from various aspects. Our innovation of modeling trustees who take the lead in fostering trust opens a new door for tackling the social dilemma. Along the direction of trustees' taking initiatives, one can investigate other approaches trustees are likely to employ to boost trust, which we leave for future study.

References

Alonso, R. and O. Câmara (2016a): "Persuading Voters," *American Economic Review*, 106, 3590–3605.

——— (2016b): "Political disagreement and information in elections," *Games and Economic Behavior*, 100, 390 – 412.

Andreoni, J. (2018): "Satisfaction Guaranteed: When Moral Hazard Meets Moral Preferences," *American Economic Journal: Microeconomics*, 10, 159–89.

Arruñada, B., M. Casari, and F. Pancotto (2015): "Pro-sociality and strategic reasoning in economic decisions," *Frontiers in Behavioral Neuroscience*, 9, 140.

Asriyan, V., D. Foarta, and V. Vanasco (2019): "Social Capital as Good Culture," Stanford University Graduate School of Business Research Paper 19-6.

Athey, S. and J. Levin (2018): "The value of information in monotone decision problems," *Research in Economics*, 72, 101 – 116.

Baader, M. and A. Vostroknutov (2017): "Interaction of reasoning ability and distributional preferences in a social dilemma," *Journal of Economic Behavior & Organization*, 142, 79–91.

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

Bergemann, D. and S. Morris (2019): "Information Design: A Unified Perspective," *Journal of Economic Literature*, 57, 44–95.

Blackwell, D. (1951): "Comparison of Experiments," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, Calif.: University of California Press, 93–102.

——— (1953): "Equivalent Comparisons of Experiments," *Ann. Math. Statist.*, 24, 265–272.

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

Bolton, G. E., E. Katok, and A. Ockenfels (2004): "How Effective Are Electronic Reputation Mechanisms? An Experimental Investigation," *Management Science*, 50, 1587–1602.

Bracht, J. and N. Feltovich (2009): "Whatever you say, your reputation precedes you: Observation and cheap talk in the trust game," *Journal of Public Economics*, 93, 1036–1044.

Brown, M., A. Falk, and E. Fehr (2004): "Relational Contracts and the Nature of Market Interactions," *Econometrica*, 72, 747–780.

Burks, S., J. Carpenter, and E. Verhoogen (2003): "Playing both roles in the trust game," *Journal of Economic Behavior & Organization*, 51, 195–216.

Cabrales, A., O. Gossner, and R. Serrano (2013): "Entropy and the Value of Information for Investors," *American Economic Review*, 103, 360–377.

Camerer, C. F., T. Ho, and J.-K. Chong (2004): "A Cognitive Hierarchy Model of

Games," *The Quarterly Journal of Economics*, 119, 861–898.
Charness, G., N. Du, and C.-L. Yang (2011): "Trust and trustworthiness reputations in an investment game," *Games and Economic Behavior*, 72, 361–375.
Charness, G. and M. Dufwenberg (2006): "Promises and Partnership," *Econometrica*, 74, 1579–1601.
Chen, C.-C., I.-M. Chiu, J. Smith, and T. Yamada (2013): "Too smart to be selfish? Measures of cognitive ability, social preferences, and consistency," *Journal of Economic Behavior & Organization*, 90, 112–122.
Coleman, J. (1990): *Foundations of Social Theory*, The Belknap Press of Harvard University Press.
Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2013): "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications," *Journal of Economic Literature*, 51, 5–62.
Crawford, V. P. and N. Iriberri (2007a): "Fatal Attraction: Salience, Naïveté, and Sophistication in Experimental "Hide-and-Seek" Games," *American Economic Review*, 97, 1731–1750.
——— (2007b): "Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?" *Econometrica*, 75, 1721–1770.
Fehr, E. (2009): "On The Economics and Biology of Trust," *Journal of the European Economic Association*, 7, 235–266.
Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10, 171–178.
Ganuza, J.-J. and J. S. Penalva (2010): "Signal Orderings Based on Dispersion and the Supply of Private Information in Auctions," *Econometrica*, 78, 1007–1030.
Glaeser, E. L., D. Laibson, J. Scheinkman, and C. L. Soutter (2000): "Measuring Trust," *The Quarterly Journal of Economics*, 115, 811–846.
Guiso, L., P. Sapienza, and L. Zingales (2004): "The Role of Social Capital in Financial Development," *American Economic Review*, 94, 526–556.
——— (2006): "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives*, 20, 23–48.
——— (2008): "Social Capital as Good Culture," CEPR Discussion Papers 6657, Centre for Economic Policy Research.
——— (2009): "Cultural Biases in Economic Exchange?" *The Quarterly Journal of Economics*, 124, 1095–1131.
Hedlund, J. (2017): "Bayesian persuasion by a privately informed sender," *Journal of Economic Theory*, 167, 229–268.
Huck, S., G. K. Lünser, and J.-R. Tyran (2012): "Competition fosters trust," *Games and Economic Behavior*, 76, 195 – 209.
Jones, G. (2008): "Are smarter groups more cooperative? Evidence from prisoner's dilemma experiments, 1959–2003," *Journal of Economic Behavior & Organization*, 68, 489 – 497.

Kamenica, E. (2019): "Bayesian Persuasion and Information Design," *Annual Review of Economics*, 11, 249–272.

Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

Karlan, D. S. (2005): "Using Experimental Economics to Measure Social Capital and Predict Financial Decisions," *American Economic Review*, 95, 1688–1699.

Knack, S. and P. Keefer (1997): "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation," *The Quarterly Journal of Economics*, 112, 1251–1288.

Lehmann, E. L. (1988): "Comparing Location Experiments," *Ann. Statist.*, 16, 521–533.

Masuda, N. and M. Nakamura (2012): "Coevolution of Trustful Buyers and Cooperative Sellers in the Trust Game," *PloS one*, 7, e44169.

Nagel, R. (1995): "Unraveling in Guessing Games: An Experimental Study," *American Economic Review*, 85, 1313–1326.

Ortoleva, P. (2012): "Modeling the Change of Paradigm: Non-Bayesian Reactions to Unexpected News," *American Economic Review*, 102, 2410–36.

Osborne, M. J. and A. Rubinstein (1994): *A Course in Game Theory*, vol. 1 of *MIT Press Books*, The MIT Press.

Perez-Richet, E. (2014): "Interim Bayesian Persuasion: First Steps," *American Economic Review*, 104, 469–74.

Persico, N. (2000): "Information Acquisition in Auctions," *Econometrica*, 68, 135–148.

Stahl, D. O. and P. W. Wilson (1994): "Experimental evidence on players' models of other players," *Journal of Economic Behavior & Organization*, 25, 309 – 327.

——— (1995): "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10, 218–254.

Zak, P. J. and S. Knack (2001): "Trust and Growth," *The Economic Journal*, 111, 295–321.