Male, Female, or No Comment?
Gender Information Disclosure in Trusting and Reciprocating

Yunwen He
Tsinghua University

Jaimie W. Lien
The Chinese University of Hong Kong

Jie Zheng
Tsinghua University

Current Version: July 30, 2021
Initial Version: November 30, 2020

Abstract:[1]

When interacting with others, individuals are often known to adjust their behavior based on the gender characteristics of the other person. Information about another person's gender tends to influence both behavior towards that individual, as well as expectations about that individual's behavior in return. However, as many societies around the world become increasingly interested in gender equality, what are the potential effects of introducing a gender-blind option? In this study, we examine the effect of gender and gender information disclosure, on decisions about giving and reciprocating, in a laboratory experiment. Treatments vary by the type of reciprocity examined (*direct*, *indirect*) and information conditions (*no information*, *imposed information*, *self-disclosed information*). Direct reciprocity combined with imposed gender information generates the highest initial offers. In addition, we find that choosing to conceal one's own gender information in the self-disclosed condition was penalized by peers via lower first stage offers from both males and females. Experience with the game generally widens the gender gap in offers made, even though players are largely similar in their level of trustworthiness. Finally, we find evidence that subjects in the self-disclosure condition attempt to reveal or conceal their gender information strategically, exhibiting experimentation following a low offer received. These findings contain implications for the design of gender information policies in various settings.

Keywords: Trust, Reciprocity, Gender, Information Disclosure
JEL codes: J16, C91, D83, D91

---

[1] He: Department of Economics, School of Economics and Management, Tsinghua University (heyw.18@sem.tsinghua.edu.cn); Lien: Department of Decision Sciences and Managerial Economics, CUHK Business School, The Chinese University of Hong Kong (jaimie.academic@gmail.com); Zheng: Department of Economics, School of Economics and Management, Tsinghua University (jie.academic@gmail.com)
We are grateful to Zihui Wang for programming assistance. For helpful comments, we thank Yvonne Chen, Yang Song, Xianghong Wang, Jubo Yan, Yang Yang, participants in 2021 China Meeting of the Econometric Society, 2021 Annual Conference of the International Association for Research in Economic Psychology and the Society for the Advancement of Behavioral Economics (IAREP-SABE), 2021 Tsinghua Platform for Microeconomics: Empirics, Experiment, and Theory, and seminar participants at Shandong University. We gratefully acknowledge research funding support from the National Natural Science Foundation of China (72073080, 7187307), Hong Kong Research Grants Council (14501919, 14500516), and Chinese University of Hong Kong Direct Grants.

Encouraging gender equality is one of the frequently discussed social policy objectives around the world. The reality in most societies is that individuals of different and same genders need to frequently interact with one another in the workplace, the marketplace, and other relevant settings in which economic surplus has the potential to be created. In traditional in-person economic exchanges, an individual's gender information is often automatically discerned by the interacting party. However, with the increasing prevalence of online interactions, gender information is not only becoming less obvious than before, but individuals can increasingly decide whether to reveal their gender or not. Indeed, studies have shown that requiring gender-blind evaluation in some settings can improve women's outcomes (Goldin and Rouse, 2020).

Given the history and path dependence of gender stereotypes and different social expectations based on gender, how do gender information policies affect the treatment of males, females and those of undeclared gender, by males and females respectively? Such questions are especially relevant in modern societies, in which flexible gender roles and gender identity concepts are increasingly recognized.[2] In this study, we seek to understand gender interaction effects in a surplus-generating trust and reciprocity game in a laboratory experiment, under differing policies for gender information provision.

Variation in gender information conditions in the real world can be seen through several natural examples in online-related settings. The case of no gender information applies readily to a fully online situation in which neither party has the opportunity or means to reveal or discern gender. Imposed or required gender information corresponds to most in-person situations in which interacting parties can infer the other party's gender, regardless of whether that party would like to disclose it or not. Other situations may be categorized into what we consider a self-disclosure setting. In many contexts one or more parties have an option to reveal their gender to the other party, through some features of online interaction, such as email signature or screen-name.[3] What are the consequences of gender self-disclosure, given decision-makers' possible attitudes towards individuals of different genders and their disclosure choices?

Trusting and reciprocating behavior has been widely studied in the laboratory environment by experimental economists using the framework of Trust games (Berg et al., 1995). Subjects playing the role of senders decide to share part of their endowment, with the tripled amount entering the pocket of responders. The responders then choose to return as much or as little of the money they have available, to the senders. The socially optimal result has a sender fully trusting the responder to reciprocate, thus transferring all his available funds. However, a purely self-interested subject would send nothing as a sender and furthermore return nothing as a responder.

However, a consensus has not yet been reached in terms of the gender effects on sender behavior and responder behavior. Some early works unrelated to the issue of gender information provision show that men are equally trusting as (Croson & Buchan, 1999; Cox & Deck, 2006) or more trusting (Snijders & Keren, 2001; Chaudhuri & Gangadharan, 2007; Migheli, 2007) than women, and women are equally trustworthy as (Cox & Deck, 2006; Migheli, 2007) or more trustworthy (Croson & Buchan,

[2] In terms of gender roles, most modern societies have expanded the range of home and labor market responsibilities acceptable to each gender. In terms of gender identity, whereas earlier gender classification systems relied on biological definitions of gender, self-identified gender is also increasingly considered.

[3] For example, the academic submission system editorialexpress.com has a gender information question with the options "Male", "Female", "Other" or "leave blank (unanswered)".

1999; Snijders & Keren, 2001; Chaudhuri & Gangadharan, 2007) than men. However, a very few studies find that women send more funds (e.g. Bohnet et al., 2010) and men return a higher proportion of money (e.g. Bellemare and Kröger, 2007; Dittrich, 2015).[4] Some studies have found that other demographic factors such as age, can interact with gender effects (Garbarino & Slonim, 2009). Comprehensive literature reviews can be found in Croson & Gneezy (2009) and Eckel & Grossman (2008). In a field study on physician referrals, Zeltzer (2020) finds evidence for same-gender bias, which in a setting of greater proportion of referring males, disadvantages female physicians.[5]

These basic results about gender differences in reciprocity games are enriched through the literature on potential provision of information regarding the gender of players. By announcing the gender-specific first names of their counterparts, Buchan et al. (2008) find that neither gender is trusted more and neither gender is reciprocated more. However, Eckel & Wilson (2003) show that women are less likely to be trusted than men in written information format, but are equally trusted in photographic format. Several studies find that males or females show a bias towards partners of the opposite gender under different scenarios (Solnick, 2001; Eckel & Wilson, 2005; Dufwenberg & Muren, 2006; Innocenti & Pazienza, 2006; Schwieren & Sutter, 2008), including the partner selection process (Slonim & Guillen, 2010). Ben-Ner et al. (2004) also show that information provision can induce women to give less to women than to men and paired members of unknown gender.

One caveat is that the underlying gender differences in trust and trustworthiness could be overestimated on the basis of one-off play. For example, when the Trust game is played for multiple periods, Chaudhuri & Sbai (2011) and Chaudhuri et al. (2013) find that although there are indeed early differences in gender, they do not persist over time. Such learning effects are also found in the repeated ultimatum game studied by Eckel & Grossman (2001) and McGee & Constantinides (2013). Therefore, the effect of repeated play is potentially important, and is one feature of our current study. In prior studies, gender information is nearly always exogenously imposed by the experiment without differentiating between the potential effects of mandatory and voluntary disclosure.

Our paper is to our knowledge, unique to the literature in providing insights on the effect of *endogenous* gender information.[6] A recent exception is Drouvelis et al (2020) which differs from our work in allowing the possibility for participants to misrepresent their gender, and examining the effect on subsequent cooperation. In this sense, the focus of their study is mainly on the impact of false information (utilizing gender as a manipulable feature), whereas our study is focused on the effects of gender and actual gender information. We complement the earlier studies which focused on Trust games, by allowing responders to reveal their gender to the paired senders in some treatments, while preserving the anonymity of interaction. We prohibit dishonest reporting in our setting, by limiting to a binary disclosure decision for responders: they can only decide whether or not to disclose, and their gender will then be truthfully revealed by the experimental program, conditional on disclosure. Furthermore, we utilize the strategy method to collect data on sender decisions, that is, senders have to choose a response under each possible situation given by the recipient's self-disclosure decision and

[4] At a group rather than individual level, studies find either no significant difference in behavior between single-sex groups and individuals of the same sex or between single-sex and mixed-sex groups (Cox, 2002; Chaudhuri et al. 2013).

[5] Our experiment deliberately recruits male and female subjects in equal numbers, so a proportion-based channel for unequal gender outcomes can be ruled out in our setting.

[6] In terms of experimental studies on voluntary versus mandatory information disclosure, a relevant reference is Kamei (2020), albeit not in the domain of gender information. Studies on gender information disclosure choices, although mainly apart from the experimental economics context, include McAllister (1980) and Derlega et al (1981).

real gender. In the other two information conditions, we either provide no information regarding the responders' gender, or automatically inform senders of the gender of their partners after they have made the choices with strategy method.

Our study can also be placed within the broader literature on information influences in social preference related games. Small and Loewenstein (2003) and Stanca et al. (2009) manipulate the timing of information provision: whether players have knowledge about the relevant information before or after their decision-making. Individuals tend to be kinder to a specific target than an unknown partner *ex ante* (Small & Loewenstein, 2003). In addition, responders reciprocate significantly more when their donors' action abstains from strategic motivation and is conducted before their awareness of potential reciprocating opportunities (Stanca et al., 2009). Servátka (2009) varies the content of information provision in a two-stage game with role reversal, and disentangles the explanatory power of reputation from social influence on the prosocial behavior. However, none of these studies examine gender differences. Thus, our study plays a key role in bridging the gap between information disclosure and gender effects.

Another key dimension of focus in our study is the type of reciprocity setting. One of our motivations is to better understand the potential differences in social interactions related to gender that are introduced via modern online and other shorter-term interactions. Indirect reciprocity emerges when the reciprocal act is directed to an unrelated third party and cannot be materially returned to the original giver, which can be understood as "You help me and I will help someone else" (Nowak & Sigmund (2005)). While direct reciprocity ("You help me and I will help you") has been primarily at the center of attention in the literature, knowledge about indirect reciprocity is sparser. Nevertheless, upstream indirect reciprocity is appropriate for our understanding of how gender stereotypes may develop, in other words, how beliefs about gender characteristics are shaped by potential reactions to previous unrelated interactions.[7] Some relevant prior research studying various types of indirect reciprocity in the laboratory and in field environments include Greiner and Levati (2005), Engelmann & Fischbacher (2009), Riyanto & Zhang (2014), van Apeldoorn and Schram (2016), and Mujcic & Leibbrandt (2017).

Relatedly, our work also contributes to a growing literature which compares the relative strengths of direct reciprocity and indirect reciprocity. Evidence on which type of reciprocity is stronger is divided. Although studies from the Trust game show that direct reciprocity can dominate (Buchan et al., 2002) or be equally intense as indirect reciprocity (Dufwenberg et al., 2001; Li, 2018), responders are found to be more rewarding in the case of indirect reciprocity in a symmetric version of the gift-exchange game (Stanca, 2009). The reciprocal behavior towards a third party is no weaker than to the original donor in a two-part dictator game where the interaction is zero-sum in nature (Ben-Ner et al., 2004; Herne et al., 2013).

Given the scarcity of prior studies on direct versus indirect reciprocity more generally, and based

[7] Upstream indirect reciprocity, as we focus on in this study, contrasts with the 'downstream' version in that the direction of giving is purely in a single direction, without any material incentive to be generous (A helps B, and B later helps C). In downstream indirect reciprocity, knowing the structure of the potential reciprocity, the original sender may anticipate the possibility to be repaid by a third party, and thus behave generously initially (B helps C, and A later helps B). Okada (2020) provides a useful summary. Due to this key difference between the two types of indirect reciprocity, we focus on the upstream type because it is in some regards, even more indirect than the downstream type (no chance for any sender to be paid back in this specific set of interactions), and thus better suited to measure the implications of personal experiences with particular demographic groups on social preferences towards different groups more broadly.

on the potential interaction of this feature with gender information, our experiment further varies the two types of reciprocity, which gives us six treatments in total. We follow Stanca (2009) in choosing the symmetric gift-exchange game to highlight the ex-ante symmetric roles of players. In addition, subjects play the game for 10 periods with random partner re-matching in each period to avoid effects from repeated interactions between the same individuals. The main purpose of our paper is to capture the general and interaction effects of gender and gender information. By varying the reciprocity type and gender information conditions on the responder side while learning and experience are taken into account, we can draw insights on social learning and interactions based on gender effects.

To summarize, our main contributions are twofold. Firstly, we fill in a gap in the experimental literature by testing the effects of information provision rules on trust and reciprocation. This allows us to answer questions of whether senders behave differently when gender information of their counterparts is exogenously given or disclosed by themselves, and whether responders' reciprocal behavior is affected by their self-disclosure decisions. Secondly, for the case of self-disclosure of gender information, we engage in a deeper investigation into the gender differences in how individuals trust others and are trusted. Our study is thus well-positioned to contribute to predictions on behavioral responses to different gender disclosure policies, and to understand the underlying motives.

To preview our main findings, we find that the greatest initial offers are generated under direct reciprocity with the externally imposed gender information condition. In terms of overall differences in sender behavior by gender, males generally give higher offers, driven by the direct reciprocity treatments. Notably, both male and female senders showed an aversion to trusting partners with non-disclosed gender information by giving lower offers. This finding raises some potential concerns for mechanisms which aim to achieve greater gender equality by making the declaration of gender optional, which is that non-disclosure has a risk of being interpreted negatively by some individuals.

Examining the dynamic feature of the data, experience gained over several periods tends to widen the gender gap in behavior, while males, females and undisclosed-gender individuals tend to exhibit similar degrees of aggregate reciprocity, albeit in different scenarios. Finally, we find evidence that gender self-disclosure choices seem to be at least partially strategic in nature, and further that experimentation in disclosure choice tends to respond to individuals' previous offer outcomes.

The remainder of the paper is organized as follows. Section 2 describes our experimental design and procedures, and Section 3 presents the experimental results, including findings based on reciprocity type, gender effects, and information treatments, by initial offer and reciprocal offers, and finally dynamic effects. Section 4 summarizes and discusses.

2. Experimental design

2.1 Treatments

We use a 2×3 design in which the six treatments differ by type of reciprocity tested (direct reciprocity (DR) or indirect reciprocity (IR)) and the information condition for the responder's gender (no information (n), imposed information (i) or self-disclosed information (or chosen information, c)). We choose a repeated play design to better understand the potential dynamic patterns of trusting and reciprocating behavior by the decision-maker (Chaudhuri et al., 2013). The decision-making part of the experiment consists of 10 periods. At the beginning of each session, subjects are randomly assigned to the role of either the sender (referred to as "player 1") or the responder ("player 2"), which remains unchanged over the course of the experiment.

The basic game is a symmetric version of the gift-exchange game (Stanca et al., 2009), which involves two stages in each period. In the first stage, player 1 is endowed with 10 tokens and decides on the amount (integer between 0 and 10) to send to player 2. This amount is subtracted from the payoff of player 1, and after being tripled, added to the payoff of player 2. In the second stage, player 2 is endowed with 10 tokens and decides on the amount (integer between 0 and 10) to send to player 1. Again, this amount is subtracted from the payoff of player 2, and is tripled before reaching player 1. Thus, each player will have a chance to play the role of sender as well as receiver. At the end of each stage, both players observe the decisions made by themselves and their paired member, and their account balances on the screen. The total balance is the summation of the payoffs in the two stages. To avoid potential gender differences under time pressure, unrestricted time was allowed for making each decision (ex. Xie, Page and Hardy, 2017).

In all treatments, paired members are randomly re-matched at the beginning of each In addition, for the indirect reciprocity treatments, subjects are additionally re-matched at the beginning of the second stage within a given period of play. Thus, in the direct reciprocity treatment, from the responders' perspective, the recipient of the reciprocal act is the same person from whom the trusting act comes from, while under the indirect reciprocity treatment, the recipient of the reciprocal act is a different person. Thus, in accordance with the definitions in the literature, the direct reciprocity treatments represent situations of repeat interaction between the same two individuals, whereas the indirect reciprocity treatments represent a generalized reciprocity to a random individual, which can be interpreted as indicative of more general social attitudes and preferences.

Subjects in the no information treatment play the standard game as we have described above. However, in the chosen information treatment, there is a pre-stage before the first stage, in which player 2 has to choose whether to make his gender known to his partner within the period. If player 2 chooses no, player 2's gender will be unknown to player 1; if yes is chosen, then player 1 will learn about player 2's gender before deciding how many tokens to send.

In both the imposed and self-disclosed information treatments, we apply a variant of the strategy method (SM) on the sender side in order to obtain more observations and gain a more complete picture of subjects' gender information-based strategies across different scenarios. Specifically, player 1 must give a response for each possible gender (male or female) of player 2 before being informed of the actual gender in the imposed information treatment, and for each possible gender information (unknown, male or female) in the self-disclosed information treatment. The actual action of player 1 is determined by the paired member's actual gender and self-disclosure decision, which is shown to both players at the end of the first stage. Table 1 summarizes the treatments in our study.

Table 1. Treatment Overview

Treatment	*Pre-stage*	*First stage*	*Re-matching*	*Second stage*
	Decision-maker	Decision-maker (elicitation method)		Decision-maker (elicitation method)
DRn	-	Player 1 (DM)	N	Player 2 (DM)
DRi	-	Player 1 (SM)	N	Player 2 (DM)
DRc	Player 2	Player 1 (SM)	N	Player 2 (DM)
IRn	-	Player 1 (DM)	Y	Player 2 (DM)
IRi	-	Player 1 (SM)	Y	Player 2 (DM)
IRc	Player 2	Player 1 (SM)	Y	Player 2 (DM)

Note: DM refers to the decision method, and SM refers to the strategy method.

2.2 Experimental procedure

Six experimental sessions in total, each session corresponding to one treatment of 26 subjects (13 males and 13 females), were conducted on October 10th and 11th, 2020 at the Economic Science and Policy Experimental Laboratory (ESPEL) of Tsinghua University, with university undergraduate students as the subject pool. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).

The gender of subjects was initially assessed based on experimenters' observation, and later verified using the post-experiment survey self-reported question on gender, which had two options: male or female. Subjects' national ID numbers were also collected, which contain their official government gender information, and can be cross-checked to the self-reported survey answers. In each session, among these three different sources of subjects' gender information, there was no case of discrepancy.

At the beginning of each session, male subjects were assigned to be seated in front of computers numbered 1-17, and female subjects were seated in front of computers numbered 18-34. The computer terminals were isolated via partitions, and any verbal communication between subjects was not allowed.

Subjects received a copy of the experiment instructions and were instructed to understand the task through the instructions provided. A translated copy of the experiment instructions is provided in the Appendix. Subjects were well-informed that there would be two stages in the game in which each role (player 1, player 2) took turns to make offers. In addition, we added some general instructions in the imposed and self-disclosed information treatments to explain the way that we obtained their gender, and to emphasize that males and females were recruited to the session equal numbers so as to hold constant subjects' neutral beliefs about the gender composition of the session.

To eliminate possible income effects, earnings were randomly selected from one out of the ten periods with the conversion rate 1 token = 1 CNY, and subjects were informed as such in the instructions. The experiment did not begin until everyone confirmed to understand the experiment rules and determination of earnings. The payment received averaged 43.43 CNY per subject (including

a 10 CNY show-up fee) for a session lasting approximately 30 minutes.[8]

3. Results

We present the experimental results with respect to the first stage offer and the second stage offer in sequence. We proceed with the analyses in several steps. We first investigate how senders behave in each of the six treatments, separately for males and females. We then discuss the gender interaction effect. In the third step, we explore the time dynamics of sender behavior. We then conduct similar analyses on the responder side, and focusing mainly on the gender differences in trustworthiness, as well as the determinants of amount returned and the gender self-disclosure decisions.

3.1 First stage offers

We begin by comparing sender behavior under the different reciprocity treatments and information conditions. Table 2 displays our basic findings, in which the statistical significance of differences in sender behavior by gender are displayed. In the following subsections, we discuss the differences in the amount sent to the responder based on experimental treatment variables (reciprocity types and information conditions), detailed gender differences of the sender, and detailed gender differences of the sender in the chosen information treatments.

3.1.1 Reciprocity Type

Firstly, we note that offers in the direct reciprocity treatments are nearly three times those in the indirect reciprocity treatments (Wilcoxon rank-sum test, $p < 0.01$ for all pairwise comparisons). Senders pass over half of their endowments to the responders who have the opportunity to directly reciprocate, while the average amount sent is only around 1.5 tokens to the responders in the indirect reciprocity condition. This indicates that sender behavior is to large extent strategically motivated, since the senders may expect to earn more through reciprocal cooperation with the same responder, but such incentive is eliminated when they will not meet the same partner again in the indirect reciprocity treatments. The result is summarized as follows:

Result 1 (Reciprocity Type and Trusting): *Direct reciprocity generates significantly higher levels of trusting than indirect reciprocity does.*

3.1.2 Gender Information Conditions

Regardless of whether in the direct or indirect reciprocity treatments, offers are the largest under the imposed gender information about responders (Wilcoxon rank-sum test, DRi vs DRn: $p = 0.027$; DRi vs DRc: $p < 0.001$; IRi vs IRn: $p = 0.037$; IRi vs IRc: $p = 0.130$), which suggests that imposed information provision has the best potential to build trust. By comparison, responders in the chosen information condition receive slightly less compared to the benchmark treatment with no information, although the difference is statistically insignificant (Wilcoxon rank-sum test, $p=0.513$). This leads to our second main result as follows:

[8] Note that every session also included another 10 periods of reciprocity games in which the subjects switched roles (player 1 versus player 2). Subjects were not informed of ahead of time about the total sequence of games played in one session, and thus the second set of 10 periods has no effect on subjects' behavior in the first 10 periods of games analyzed in this study. The average payment for the first 10 periods reported in this study is 27.05 CNY per subject including the show-up fee, which took in total approximately 20 minutes.

Result 2 (Information Condition and Trusting): *Imposed gender information about the responder leads senders to send more than in either the no information or chosen information treatments.*

3.1.3 Results by Gender

Next, we consider how the amount sent varies with the gender of player 1. Splitting the data in the male senders and female senders reveals that men are substantially more generous than women in the direct reciprocity treatments, while the gender gap narrows in the indirect reciprocity treatments and even reverses in the indirect reciprocity, chosen information treatment (IRc), in which women send significantly higher offers than men. This result corroborates many earlier findings suggesting that males are more trusting than females, and the comparison between direct and indirect reciprocity provides some support for the interpretation that males express trust mainly for an instrumental purpose (Buchan et al., 2008). [9]

Table 2. Average first stage offer

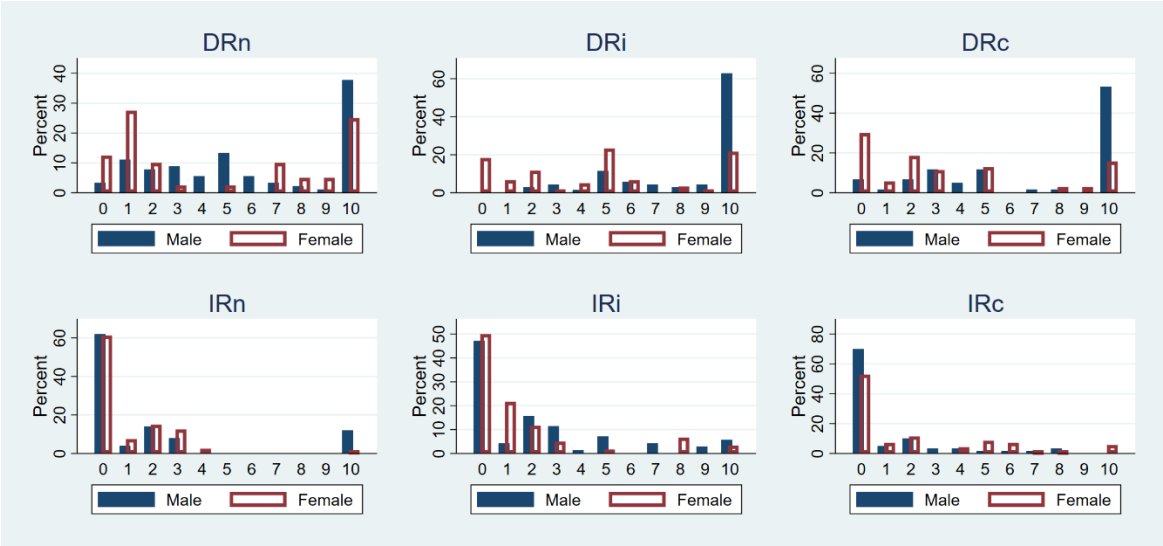
Avg. (std. dev.)	*Pooled outcome*	*Male*				*Female*			
		Outcome	Strategy method			Outcome	Strategy method		
			Unknown	To male	To female		Unknown	To male	To female
DRn	5.638	6.044				4.725**			
	(3.777)	(3.576)				(4.095)			
DRi	6.669	8.357		7.671	8.843	4.7***		4.683***	4.967***
	(3.547)	(2.479)		(3.025)	(2.103)	(3.605)		(3.652)	(3.556)
DRc	5.038	6.867	6.033	7.3	6.667	3.471***	2.243***	3.714***	3.471***
	(3.995)	(3.652)	(3.086)	(3.466)	(3.648)	(3.610)	(3.351)	(3.620)	(3.365)
IRn	1.277	1.76				0.975			
	(2.379)	(3.230)				(1.591)			
IRi	1.923	2.243		2.4	2.086	1.55		1.717	1.5
	(2.836)	(2.990)		(3.419)	(2.603)	(2.620)		(2.675)	(2.404)
IRc	1.623	1.05	1.033	1.033	1.1	2.114**	1.743*	2.243***	2.986***
	(2.642)	(2.054)	(2.058)	(2.058)	(2.039)	(2.986)	(2.477)	(2.985)	(3.479)

Notes: The Wilcoxon rank-sum tests for pairwise comparisons on the amount sent broken up *by gender* are reported in the table. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Standard deviations are presented in parentheses.

Figure 1.A, which shows the distribution of first stage offers by gender of the sender and by treatment, illustrates the detailed gender differences in trusting behavior. While both males and females were hesitant to give large offers in the indirect reciprocity treatments, males exhibited the greater distributional shift between direct and indirect reciprocity, with the modal offer shifting from the full amount to zero, respectively. Female senders comparatively speaking, exhibited a less drastic difference between the two types of reciprocity.

[9] There is no large variation in the gender distribution of senders across different treatments *ex post* (9 males and 4 females in DRn; 7 males and 6 females in DRi; 6 males and 7 females in DRc; 5 males and 8 females in IRn; 7 males and 6 females in IRi; 6 males and 7 females in IRc).

Figure 1.A: First stage offers by gender (actual data)



Note: The *x-axis* depicts the possible values of the first stage offer, and the *y-axis* depicts the percentages.

Result 3 (Gender and Trusting): *Men send significantly more than women to any type of partner under direct reciprocity, while there is no (or reverse) such gender gap in the indirect reciprocity conditions.*

3.1.4 Gender Self-Disclosure

With the previous basic results on trust behavior by treatment variables and gender established, we now examine subjects' gender revelation choices in the self-disclosure (chosen information) treatment and consequences for amounts sent.

Firstly, we find that when responders refuse to disclose their gender, they are more likely to be punished with a low offer. The average amount sent to responders that withhold their gender information is 3.714 and 1.103 tokens in DRc and IRc treatments, respectively, both smaller than the average level in the absence of information provision (Wilcoxon rank-sum test, 3.714 vs 5.638: $p = 0.054$; 1.103 vs 1.277: $p = 0.571$). Therefore, the willingness to reveal one's gender seems to play a consequential role to player 1 in evaluating the trustworthiness of player 2.

In both the chosen information treatments (DRc and IRc), in which responders are provided with the gender information disclosure choice, the amounts sent to a responder of unknown gender are significantly less than amounts sent to either male or female responders, and this is true whether the sender themselves is either male or female. For example, in the DRc treatment, males send over an average of 1 token less to unknown gender responders compared to responders who were disclosed to be male (Wilcoxon matched-pairs signed-ranks test, $p < 0.001$), and over 0.5 tokens less to unknown gender responders compared to female responders (Wilcoxon matched-pairs signed-ranks test, $p < 0.008$). For their part, female senders similarly give significantly less to unknown-gender responders, compared to responders known to be either male or female, of amounts more than 1 token less than given to either known gender (Wilcoxon matched-pairs signed-ranks tests, $p < 0.001$).

Compared to direct reciprocity, the indirect reciprocity treatment (IRc) yields less discrimination against non-disclosure of gender. For male senders, there was no difference between amounts sent to male versus unknown responders (Wilcoxon matched-pairs signed-ranks test, $p = 1$), and only a small difference in amounts sent to female versus unknown responders (Wilcoxon matched-pairs signed-

ranks test, p = 0.046). However, for female senders in the indirect reciprocity setting, the differences are of greater magnitude and significance, in terms of the gap between amounts sent to unknown gender responders and either males or females (Wilcoxon matched-pairs signed-ranks tests, $p < 0.001$). This seems to indicate that while males view gender disclosure as relatively less consequential in the indirect reciprocity setting, females tend to react more similarly to the case of direct reciprocity. This result is also consistent more generally with males being more sensitive to the distinction between direct and indirect reciprocity. Overall, women tend to be less comfortable with being uninformed, which may be supportive of previous findings on females' greater sensitivity to exploitation (Ingram & Berger, 1977).

These patterns can be seen in the distribution plots of offers between senders and responders of different types below:

Figure 1.B: First stage offer by gender, direct reciprocity (strategy data)

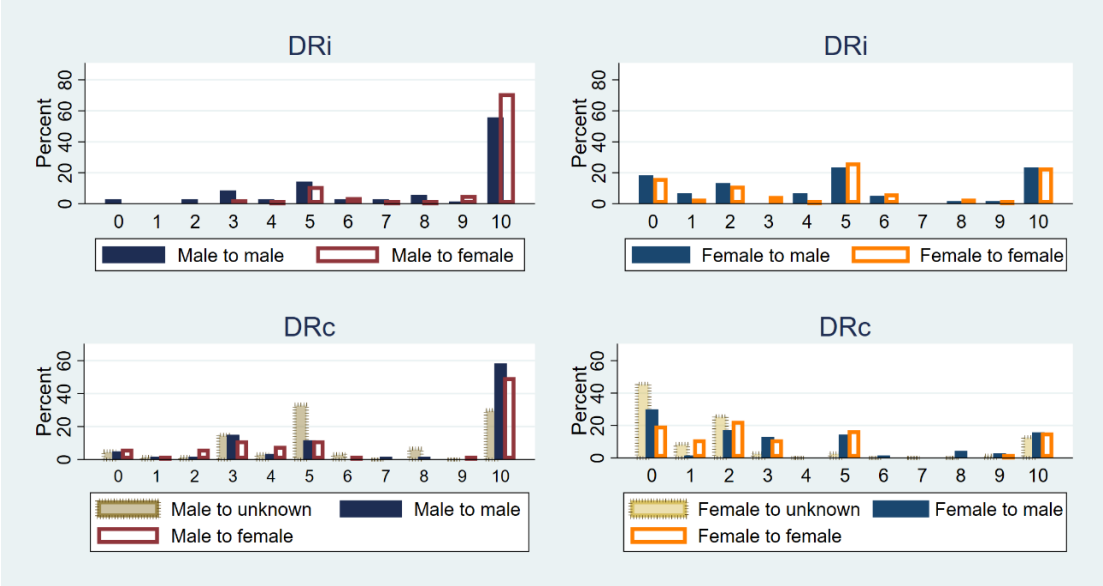
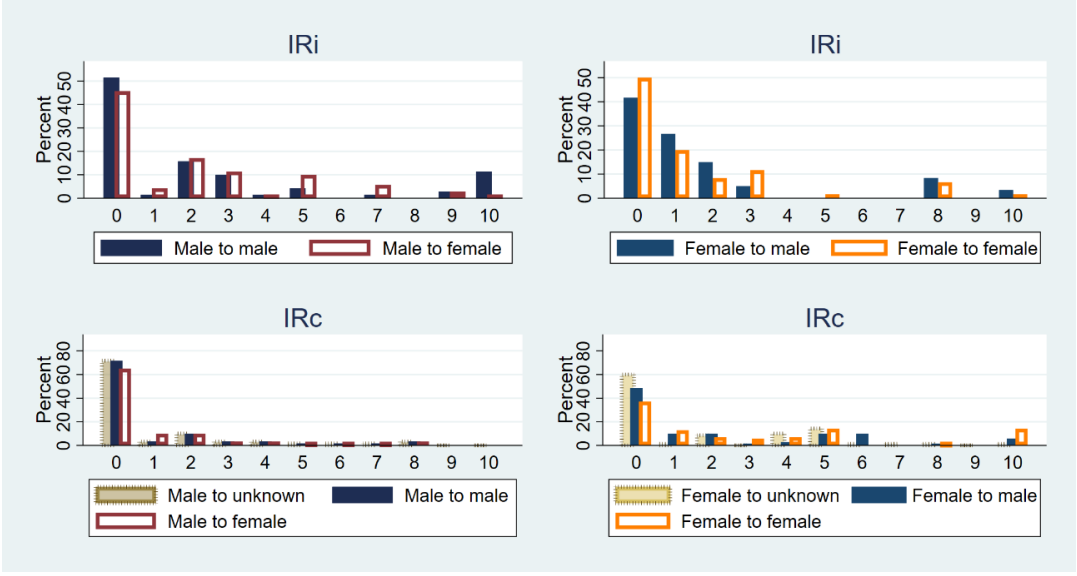


Figure 1.C: First stage offer by gender, indirect reciprocity (strategy data)



3.1.5 Regression analysis

To test and control simultaneously for the effect of the variables discussed earlier on first stage offers, we implement regression analyses with offer amount as the dependent variable. In Table 3 taking the first stage offer (*Offer1*) as the dependent variable, the treatment characteristics are controlled by a dummy variable for the indirect reciprocity condition (*Indirect*) with the direct reciprocity condition as the reference group, and another two dummy variables for imposed information (*ImpInfo*) and self-disclosed information (*ChoInfo*) conditions with no information as the reference group. *Period* is a time variable which identifies the number of repetitions of the game at that observation to control for possible time trends, and *Gender* is equal to 1 for males and 0 for females. Interaction variables are also denoted accordingly. These independent variables are used consistently across different empirical regressions in the paper.

Table 3. Tobit regression results for first stage offers

First stage offer	(1)	(2)	(3)	(4)
Period	-0.017	-0.014	-0.147	-0.168***
	(0.082)	(0.083)	(0.097)	(0.054)
Indirect	-8.393***	-8.022***	-5.275***	-1.954***
	(1.475)	(1.448)	(1.814)	(0.492)
ImpInfo	1.640	1.681	0.917	-0.129
	(1.540)	(1.474)	(1.982)	(0.560)
ChoInfo	-0.446	-0.145	0.116	-0.694
	(1.601)	(1.542)	(2.163)	(0.527)
Gender		2.715**	2.877	
		(1.264)	(2.474)	
Gender×Indirect			-5.445**	
			(2.524)	
Gender×ImpInfo			2.306	
			(2.972)	
Gender×ChoInfo			0.151	
			(3.007)	
Gender×Period			0.278*	
			(0.167)	
Lag_offer1				1.150***
				(0.099)
Lag_offer2				0.285***
				(0.072)
Constant	6.614***	4.953***	4.495**	0.233
	(1.232)	(1.449)	(2.007)	(0.546)
Observations	780	780	780	702
Left-censored	266	266	266	245
Right- censored	161	161	161	149
Pseudo R^2	0.086	0.095	0.108	0.278

Notes: Standard errors clustered at the individual level are displayed in parentheses. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

As shown in columns (1) – (3) of Table 3, player 1 gives both statistically and economically significant lower offers when their interaction with player 2 is temporary (indirect reciprocity) within a period. Although we do not find a strong impact of information provision, the coefficient on imposed information treatments remains positive, consistent with earlier results. Men send significantly more than women on average, but this phenomenon is context-specific since in the case of indirect reciprocity, they are shown to give less than women. Overall, the regression results show that our earlier results are robust to accounting for the control variables, including gender interaction effects and lagged offers.

Table 4. Tobit regression results for gender interaction effects

First stage offer	Strategy method		Actual outcome	
	DRi & IRi	DRc & IRc	DRi & IRi	DRc & IRc
Period	-0.043	-0.140	-0.079	-0.184
	(0.118)	(0.130)	(0.129)	(0.156)
Indirect	-8.463***	-6.617**	-8.541***	-7.303**
	(2.290)	(2.640)	(2.345)	(2.933)
Men-to-unknown		-1.481*		-3.190
		(0.795)		(2.272)
Men-to-women	1.000	-0.384	-0.635	-0.233
	(0.757)	(0.681)	(1.135)	(1.201)
Women-to-unknown		-4.329*		-4.601
		(2.502)		(3.239)
Women-to-men	-3.001	-2.133	-3.706*	-2.547
	(2.116)	(2.355)	(2.206)	(2.541)
Women-to-women	-3.217	-1.117	-4.906**	-2.436
	(2.172)	(2.410)	(2.407)	(2.490)
Constant	9.748***	7.540***	10.915***	8.351***
	(2.160)	(1.950)	(2.426)	(2.189)
Observations	520	780	260	260
Left-censored	146	304	74	104
Right- censored	129	129	63	47
Pseudo R^2	0.111	0.062	0.117	0.074

Notes: *Men-to-unknown* is a dummy variable equal to 1 if a male sender is matched to a responder without disclosing gender. Similar definitions apply to the dummy variables *Men-to-women, Women-to-unknown, Women-to-men, Women-to-women,* so that estimates are relative to male senders matched to male responders. Standard errors clustered at the individual level are displayed in parentheses. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table 4 shows regression results focusing on the gender effects between sender and responder. From the results we can observe that the indirect reciprocity environment remains a significant negative predictor of amount sent in the first stage. In the gender interaction terms, men-to-men serves as the comparison group. While the amount sent from men to unknown gender and from women to unknown gender are significantly lower in the strategy method, the result is not borne out in the actual outcome, likely due to low proportion of subjects actually choosing not to disclose their gender

information.[10] In addition, in the actual outcome, women gave significantly less to both men and women, although this result was not significant using the full data from the strategy method.

Overall, the results of subsections 3.1.4 and 3.1.5 lead to our result about the effect of gender non-disclosure on trust behavior:

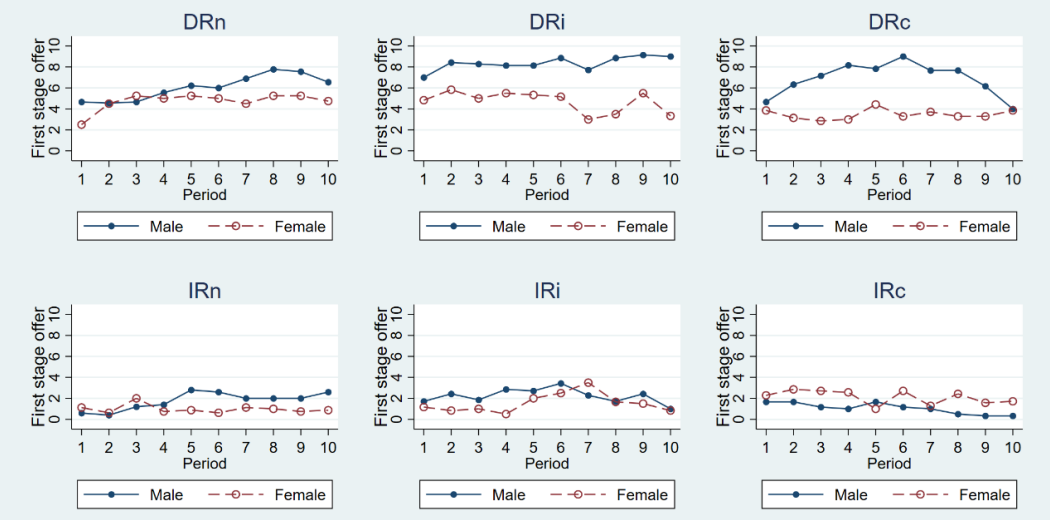
Result 4 (Effect of Gender Information Non-disclosure on Trusting): *Both men and women penalize responders that choose not to disclose their gender, by sending less compared to the case of either a known male or female responder.*

3.1.6 Time Dynamics of Offers

Our next step is to understand how the first stage offer tends to evolve over time. Recalling that in all treatments, players are randomly re-matched in each round, the time trends can be interpreted as the aggregate results of social learning through subjects' average accumulated experiences interacting with other subjects of different genders and reciprocal behaviors. The time trends might thus be useful in drawing observations about the medium-term effects of the different gender information disclosure policies.

Figure 2 illustrates the time path of first stage offers, where the *x-axis* depicts the number of periods, and the *y-axis* depicts the average actual amount sent. We can observe that the trend line of offers by males stays above that of females in most of the periods, but the distinct gap vanishes in indirect reciprocity treatments, which is consistent with the aggregate data discussed earlier. The Wilcoxon rank-sum tests on the comparisons between the first and second half (periods 1–5 and periods 6–10) of the games (see Table A1 in Appendix) indicate that the behavioral pattern is rather stable for senders, except that males exhibit a noticeable tendency to give more in DRn with the amount increasing from around 4.5 tokens in early periods to 7.5 tokens in later periods.

Figure 2. First stage offer over periods (actual data)



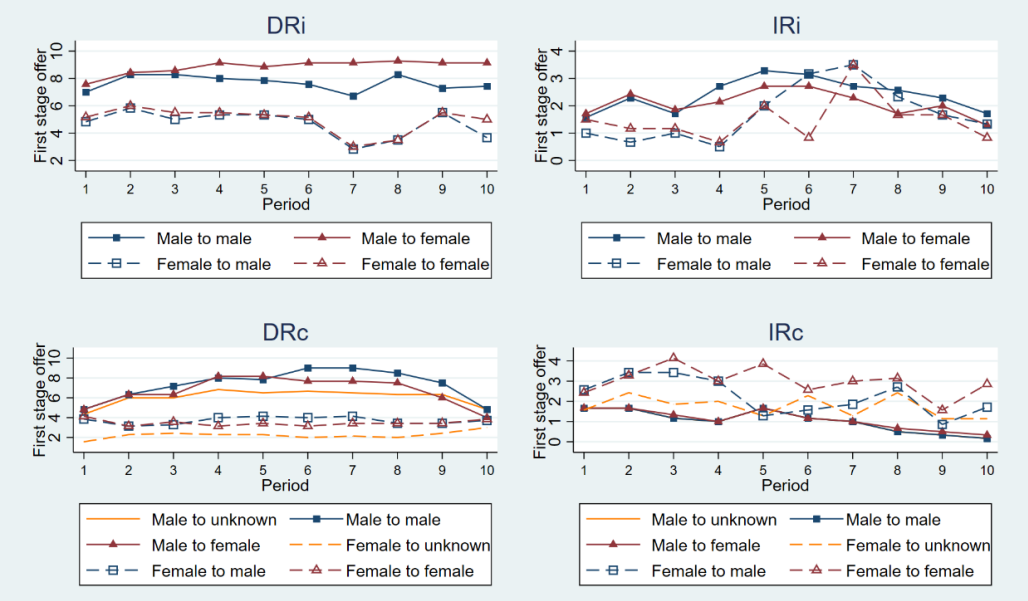
On the other hand, the gender difference seems to be relatively larger in later periods. The individual-level regression analysis reported in the third column of Table 3 indicates that the amount sent by males generally starts at a higher level and also grows at a faster rate than that of females as

[10] This also indicates the importance of the strategy method for our research design, since the gender disclosure proportions in particular are not necessarily predictable ex-ante.

far as the time trend is concerned, which contrasts Chaudhuri et al. (2013)'s findings.

Figure 3 decomposes the sent amount elicited by strategy method by responder's gender over the 10 periods. The trend lines depicting offers to male responders and to female responders are at similar levels to one another, indicating relatively low gender discrimination by senders throughout the experiment. The result is also supported by the Wilcoxon rank-sum test results in Table A1 and the regression results (not shown) including the interaction terms between period and gender pair dummies (*Men-to-women* etc.) whose coefficients are all not significant. Male responders and female responders remain nearly equally treated as time goes on.

Figure 3. First stage offer by responder's gender over periods (strategy data)



We now examine the issue of time consistency. In the column (4) of Table 3, we include the lagged variables. The estimated coefficient for the lag amount sent (*Lag_offer1*) is significantly positive and centered around unity implying that subjects' beliefs on others' trustworthiness are rather fixed. However, we still find that the reciprocal amount from the responder in the previous period (*Lag_offer2*) reshapes senders' initial beliefs, though in a smaller magnitude.

In Table 5.A, we further examine the dynamic change of player 1' s decisions on the possibly different amounts sent to males and females. Once again, several Tobit regression models are estimated in which we regress the gap between offers made to males and offers made to females (*Offergap*) on the set of observed characteristics. In Table 5.B, we consider as the dependent variable, the offer gap between disclosed and non-disclosed responders, which is calculated as the average offer made to a known male or a known female, minus the offer to a non-disclosed responder, as decided by each sender.

Subjects may update their beliefs on the trustworthiness of males and females after receiving feedback in the information treatments. Here, we consider two different sets of important history information. The first set is the average amount returned by male responders (*Avgmen*), female responders (*Avgwomen*), and those of unknown gender (*Avgunknown*) the sender had ever encountered in the past, and the second set includes the latest amount returned by male responders (*Latestmen*), female responders (*Latestwomen*) and those with unknown gender (*Latestunknown*).

In the imposed information treatments, we find that the higher the amount sent back by men in a subject's history of play, the more the subject will offer to men in the current period for given offers to women, which serves to close the originally negative offer gap or widen the originally positive offer gap. It is noteworthy that male senders seem to rely on the history information more heavily compared with female senders, given that they tend to raise the amount sent to men either with the increase in the average amount returned by male responders or with the decrease in the average amount returned by female responders. However, in the short-term, senders are mainly significantly affected by the latest amount returned by male responders with similar gender sensitivity.

A detailed examination shows that, perhaps surprisingly, senders increase offers to both genders when faced with a higher average amount returned or a higher latest amount returned by males. With the increase in the offers made for male responders being relatively larger than that for female responders, we thus observe a larger offer gap. However, senders tend to pay little attention to the reciprocity of females, and do not adjust offers to either gender quantitatively based on the history data of female responders (see also Table A3 in Appendix).

By contrast, there is only weak evidence to suggest that subjects take previous information into account when making current discriminating decisions in the chosen information treatments. The regressions show that an increase in the amount sent back by men previously in an individual's history of play, is associated with a significant increase in the offer gap between male and female responders, while having neutral or negative impact on the offer gap between disclosed and non-disclosed responders. The estimated coefficient of *Lag_offergap* is always above 0.7, indicating a strong inertia in offers.

The main robust findings can be summarized in our next result on time dynamics:

Result 5 (Time Dynamics in Trusting): *The offer gaps towards male and female responders, as well as towards gender disclosed and non-disclosed responders are generally stable over time and significantly path dependent. The gender gap for recipients tends to increase with the historical size of offers made by males.*

Table 5. Tobit regression results for offer gaps

5.A Offer gap towards male and female responders

Offer gap	DRi & IRi				DRc & IRc			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Period	-0.090	-0.083	-0.054	-0.048	-0.044	-0.025	-0.046*	-0.044*
	(0.065)	(0.066)	(0.050)	(0.048)	(0.028)	(0.023)	(0.026)	(0.025)
Indirect	0.704**	0.676**	0.883**	0.895**	-0.037	0.070	-0.126	-0.228
	(0.339)	(0.274)	(0.351)	(0.352)	(0.143)	(0.120)	(0.132)	(0.158)
Lag_offergap	0.371***	0.322***	0.382***	0.352***	0.725***	0.715***	0.726***	0.712***
	(0.110)	(0.106)	(0.112)	(0.113)	(0.081)	(0.092)	(0.076)	(0.083)
Avgunknown					0.006	0.004		
					(0.034)	(0.092)		
Avgmen	0.175***	0.092***			0.061*	0.112**		

	(0.051)	(0.033)			(0.034)	(0.053)		
Avgwomen	-0.121*	-0.026			-0.030	-0.079		
	(0.069)	(0.044)			(0.043)	(0.098)		
Gender ×Avgunknown						-0.005		
						(0.101)		
Gender ×Avgmen		0.145**				-0.083*		
		(0.069)				(0.042)		
Gender ×Avgwomen		-0.169*				0.091		
		(0.092)				(0.106)		
Latestunknown							-0.004	-0.039
							(0.021)	(0.037)
Latestmen			0.169***	0.139***			0.020	0.043
			(0.052)	(0.040)			(0.022)	(0.038)
Latestwomen			-0.060	0.015			0.013	0.020
			(0.047)	(0.033)			(0.026)	(0.045)
Gender × Latestunknown								0.064
								(0.047)
Gender ×Latestmen				0.049				-0.041
				(0.069)				(0.035)
Gender ×Latestwomen				-0.116				-0.022
				(0.073)				(0.051)
Constant	-0.073	-0.066	-0.676	-0.724*	0.117	-0.098	0.230	0.246
	(0.578)	(0.558)	(0.409)	(0.431)	(0.225)	(0.181)	(0.211)	(0.247)
Observations	174	174	174	174	109	109	112	112
Left-censored	2	2	2	2	0	0	0	0
Right- censored	0	0	0	0	0	0	0	0
Pseudo R^2	0.112	0.119	0.125	0.132	0.293	0.295	0.280	0.283

5.B Offer gap towards disclosed and non-disclosed responders (DRc & IRc)

Offer gap	(1)	(2)	(3)	(4)
Period	-0.113	-0.062	-0.110	-0.088
	(0.069)	(0.045)	(0.071)	(0.057)
Indirect	-0.089	0.318	-0.018	0.267
	(0.212)	(0.257)	(0.179)	(0.179)
Lag_offergap	0.778***	0.761***	0.787***	0.818***
	(0.122)	(0.106)	(0.102)	(0.093)
Avgunknown	-0.030	0.032		
	(0.030)	(0.058)		
Avgmen	-0.083*	0.041		
	(0.043)	(0.082)		
Avgwomen	0.080	-0.089		

	(0.073)	(0.113)		
Gender ×Avgunknown		-0.119		
		(0.114)		
Gender ×Avgmen		-0.198*		
		(0.119)		
Gender ×Avgwomen		0.303		
		(0.204)		
Latestunknown			-0.001	0.080
			(0.031)	(0.090)
Latestmen			-0.037*	-0.025
			(0.020)	(0.039)
Latestwomen			0.033	-0.026
			(0.037)	(0.026)
Gender ×Latestunknown				-0.137
				(0.099)
Gender ×Latestmen				-0.026
				(0.060)
Gender ×Latestwomen				0.138*
				(0.070)
Constant	1.258*	0.686	1.101*	0.784
	(0.745)	(0.448)	(0.656)	(0.510)
Observations	109	109	112	112
Left-censored	0	0	0	0
Right- censored	0	0	0	0
Pseudo R^2	0.275	0.290	0.263	0.278

Notes: The tables report Tobit regression results where the dependent variable in Table 5.A is the offer to a male responder minus the offer to a female responder decided by each sender, and the dependent variable in Table 5.B is the average of the offer to a disclosed male responder and a disclosed female responder minus the offer to a non-disclosed responder decided by each sender. Standard errors clustered at the individual level are displayed in parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

3.2 Second stage offers

We now examine the reciprocal behavior in the second stage across treatments. Descriptive statistics for the second stage offer are summarized in Table 6. The average amount returned is only around 2 tokens in the indirect reciprocity treatments, which is significantly lower than the general level of 5 tokens sent back in direct reciprocity treatments (Wilcoxon rank-sum test, $p < 0.05$ for all pairwise comparisons). Information also matters in the responder behavior even though they have no access to the sender's gender. There are typically significant differences in the second stage offer between the treatments with and those without gender information provision (Wilcoxon rank-sum test, DRn vs DRi: $p = 0.515$; DRn vs DRc: $p = 0.006$; IRn vs IRi: $p = 0.031$; IRn vs IRc: $p < 0.001$).

Concerning the role of gender, we either find no significant gender differences, or that men send back more than women in direct reciprocity treatments. On the other hand, male responders send back a significantly larger amount in the indirect reciprocity treatments IRn and IRi, whereas female responders are particularly generous in IRc, even returning significantly more than in DRc at the 5%

level. Investigating deeper into the phenomenon, we find that it is women who keep their gender secret that make significantly higher offers compared with the non-self-disclosed men. As a consequence, the comparisons of the average second stage offer between imposed and self-disclosed treatments point to an opposite direction under the two types of reciprocity (Wilcoxon rank-sum test, DRi vs DRc: $p = 0.003$; IRi vs IRc: $p = 0.001$).

Table 6 also indicates that females appear to be relatively more sensitive to information conditions, which corroborates some psychology evidence suggesting that women's behavior is especially susceptible to social cues (Gilligan, 1982). The non-parametric Wilcoxon rank-sum tests validate that females' behavioral patterns are not as similar within the direct or indirect reciprocity treatments (DRn vs DRi: $p = 0.259$, DRn/DRi vs DRc: $p < 0.001$; IRn vs IRi: $p = 0.084$; IRn/IRi vs IRc: $p < 0.001$) as males' (Wilcoxon rank-sum test, DRn vs DRi: $p = 0.849$; DRn vs DRc: $p = 0.466$; DRi vs DRc: $p = 0.525$; IRn vs IRi: $p = 0.060$; IRn vs IRc: $p = 0.032$; IRi vs IRc: $p = 0.848$).

Last but not least, the aggregated self-disclosure probability, though maintaining at a high level, is significantly lower in IRc than in DRc (Wilcoxon rank-sum test, 0.892 vs 0.7, $p < 0.001$), which is also because females are significantly more reluctant to reveal their gender in IRc (Wilcoxon rank-sum test, $p < 0.001$), while males' choices of disclosure does not change much between the two treatments (Wilcoxon rank-sum test, $p = 0.109$). Furthermore, the players choosing not to reveal their gender feature lower second stage offers compared with gender-disclosed subjects, but such difference could be due in part to the lower amounts they received. [11]

Table 6. Average second stage offer and self-disclosure probability

Avg. (std. dev.)	*Pooled*	*Male*				*Female*			
		Self-disclosure probability	Second stage offer			Self-disclosure probability	Second stage offer		
			Pooled	Disclosure	No disclosure		Pooled	Disclosure	No disclosure
DRn	5.108 (3.965)		4.625 (4.149)				5.322 (3.886)		
DRi	5.562 (3.892)		4.9 (4.074)				6.129 (3.663)		
DRc	4.023 (4.151)	0.829 (0.380)	5.371 (4.118)	5.603 (4.043)	4.25 (4.475)	0.967** (0.181)	2.45*** (3.629)	2.534*** (3.662)	0 (0)
IRn	1.046 (2.255)		1.325 (2.667)				0.6* (1.262)		
IRi	1.946 (3.168)		2.383 (3.528)				1.571* (2.795)		
IRc	2.977 (3.644)	0.714 (0.455)	2.557 (3.598)	3.08 (3.843)	1.25 (2.531)	0.683 (0.469)	3.467** (3.666)	3.585 (3.578)	3.211** (3.938)

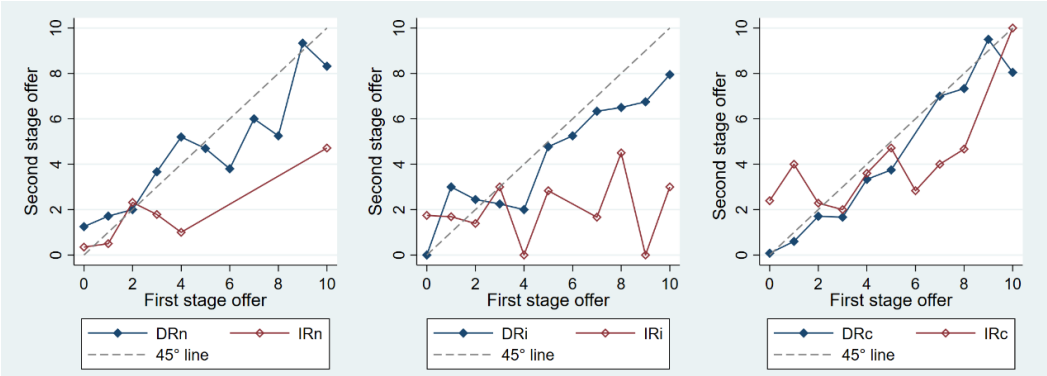
Notes: The Wilcoxon rank-sum tests for pairwise comparisons on the self-disclosure probability and amount sent broken up *by gender* are reported in the table. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Standard deviations are presented in parentheses.

[11] We also conducted another treatment DRn_2 in which two randomly-selected periods out of the ten periods would be chosen to pay. Evidence shows that the sender behavior and responder behavior are unmediated by the size of stake (first stage offer of male: 6.45 vs 6.044; of female: 5.5 vs 4.725; first stage offer of male: 5.486 vs 4.625; of female: 4.783 vs 5.322. The Wilcoxon rank-sum tests for pairwise comparisons are all not statistically significant.).

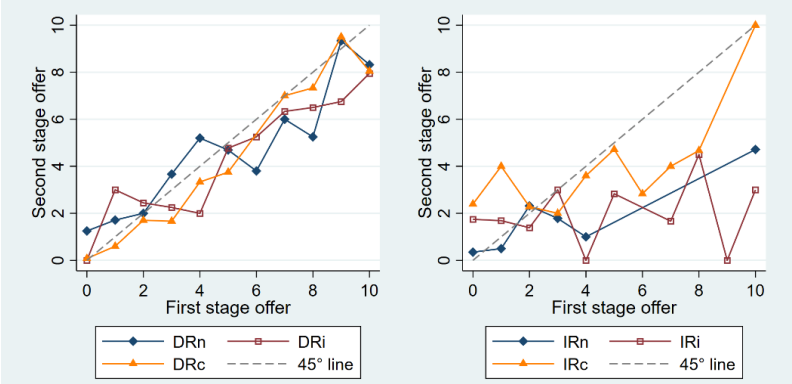
Another important insight from Table 6 is that the average second stage offer is only slightly lower than the average first stage offer, with the exception that it is even higher than the first stage offer in IRc. So far, since we have not controlled for the amount sent by player 1, we cannot evaluate player 2's propensity to reciprocate. Once the amount sent by player 1 is controlled for, we find that around 60% of responders return the same amount they received to their partners. Also, the Spearman correlation coefficients capturing the correlation between the responses of player 2 and the amount passed by player 1 are significantly positive at the 1% level in all six treatments, however, the positive relationship in the indirect reciprocity treatments is substantially weaker than in the direct reciprocity treatments.[12]

The correlograms in Figure 4 make comparisons across treatments clearer. The second stage offers are roughly aligned to each other for the three direct reciprocity treatments, and are distinctly larger than those in the corresponding indirect reciprocity treatment conditional on the first stage offer, especially when the first stage offer is high. The strength of indirect reciprocity seems to strengthen in the chosen information treatment IRc, compared with the other two information conditions. Notably, when gender information is endogenized and player 2 is entrusted with the initial counterpart's entire endowment (occurs four times in total in the data), they are also willing to pass this amount on to another re-matched stranger.

Figure 4. Responses of player 2

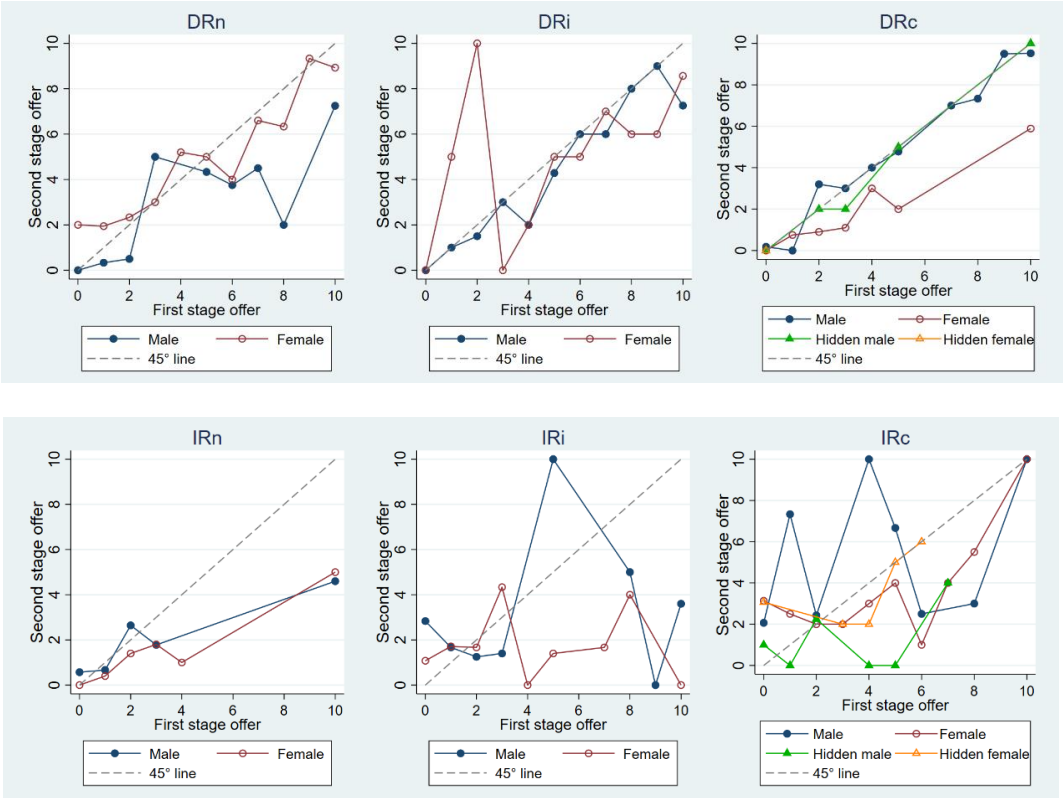


4.A Comparisons across reciprocity types



4.B Comparisons across information conditions

[12] Spearman correlations are 0.701, 0.657 and 0.737 in DRn, DRi, DRc respectively. It is reduced to 0.608 in IRn and more than halved in IRi and IRc (0.249 and 0.331 respectively).



4.C Comparisons across gender

The analogous Figure 4.C based on gender do not demonstrate any obvious differences by gender. Note that there are several outliers where women are substantially more reciprocal than men in DRi and men are substantially more reciprocal than women in IRi and IRc by returning their entire endowment for some specific values of the first stage offer. However, note that there is only one observation in each case, thus we cannot draw any conclusions about differences reciprocating trust between males and females.

Result 6 (Reciprocity Type, Information Condition, and Reciprocating): *The direct reciprocity treatments generate higher levels of reciprocity than the indirect reciprocity treatments, and there are little to no influences of information treatments, or discernable differences by gender.*

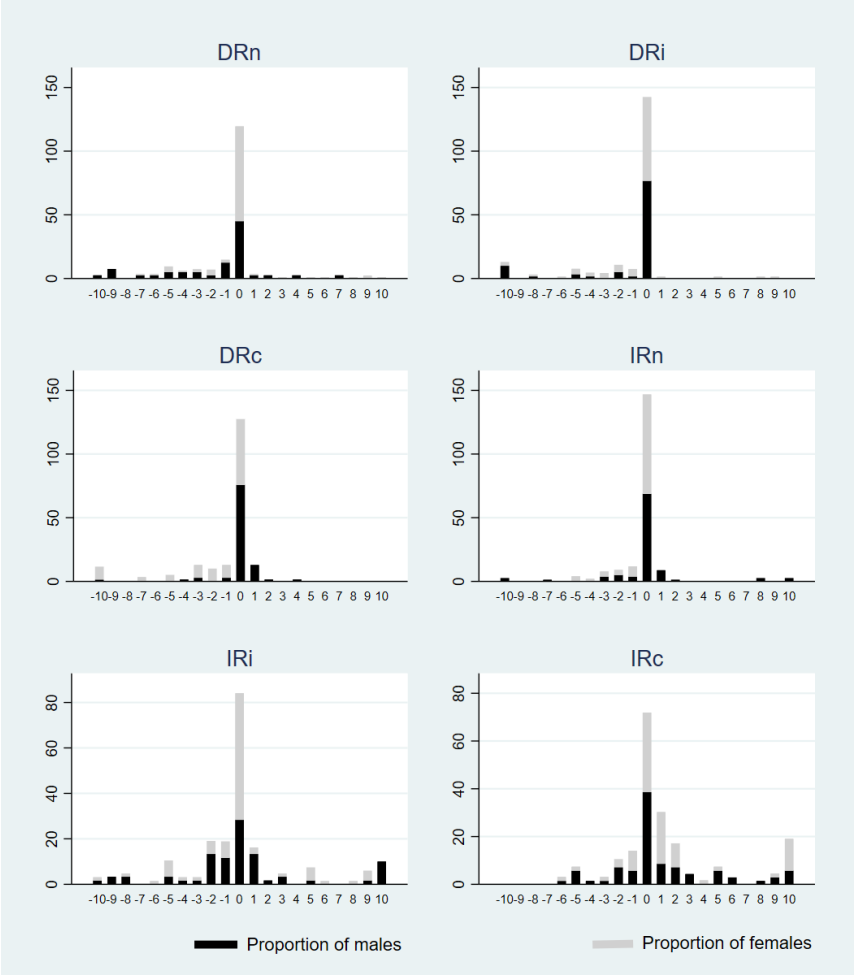
We interpret the above result cautiously since regression analysis which controls for individual observable factors does not yield statistically significant treatment effects to confirm the general trends depicted above. Controlling for individual factors, Table 7 provides a comprehensive look at the second stage offer conditional on treatment and subject characteristics. The pooled regression results shown in the first column confirms that indirect reciprocity is weaker in magnitude, but the result is not statistically significant. In addition, the difference in the degree of reciprocity is insignificant between the information treatments. Player 2's giving decisions appear to be predominantly determined by the first stage offers made to them.

In Figure 5, we depict on the *x-axis* the second stage gap between the amount sent back and received by player 2 (amount sent back minus amount received). The *y-axis* shows the proportion of male responders and female responders in each gap category, where the upper part of the bar denotes

females and the bottom part of the bar denotes males. Although the right tail of the distributions is thinner than the left tail, meaning the majority are not willing to give more than what they obtained, the vast majority display some positive degree of reciprocal behavior. A few players send back their entire endowment even if receiving nothing in the first stage. Another phenomenon that appears in our results is that the variance of the gap is relatively larger in the indirect reciprocity treatments, especially in IRi and IRc.

We find that women return relatively more compared to men under direct reciprocity, since men tend to exploit the sender. The average amount "owed" by women is 1.45 (0.70) tokens less than that of men in DRn (DRi). The pattern in DRc is reversed in that there is a possibility of 10% that women of self-disclosed gender give back nothing. This idiosyncratic behavior might be linked to the subtle discrimination towards female responders as found in Table 2. By contrast, men account for a larger proportion of the efficiency-oriented population who are willing to give more than they received under indirect reciprocity (except in IRc, 15% vs 0% in IRn, 31.67% vs 17.14% in IRi, 38.57% vs 50% in IRc). This finding is consistent with prior studies that indicate women tend to be relatively more fair and equity-minded, while men are prone to either perfect selfishness or perfect selflessness (Andreoni & Vesterlund, 2001; Miller & Ubeda, 2012).

Figure 5. Distribution of second stage gap



The data show that females are not necessarily more trustworthy than males, as the literature also

documents. Following an econometric approach, an additional variable *Disclosure* (equal to 1 for disclosed responders, and 0 otherwise) is included to control for the self-disclosure decisions in the chosen information treatments. We observe that across all the conditions presented in Table 7, gender does not add substantially to explaining the amount returned. Neither the estimated coefficients for the interaction terms between gender and treatment dummies nor between gender and the first stage offer are statistically significant at conventional levels (not shown). Thus, males and females behave similarly to each other in terms of the degree of reciprocity.

Limiting our attention to the chosen information treatments, the offers made by self-disclosed responders are relatively higher, but this difference is not close to statistically significant and not apparent by gender. We can also arrive at this finding by looking at Figure 4.C: The performance of the revealed responders and hidden responders are highly similar in DRc. The same holds for females in IRc, though the hidden males in IRc are more likely to profit at the expense of senders by always returning a low offer.

We conduct similar regression analyses to Table 7 for each treatment in greater detail (see Table A5 in Appendix). In addition, to find a single index to measure subjects' trustworthiness, we use the ratio of the second stage gap between the amount offered by the responder (*Offer2*) and the amount offered by his first-stage counterpart (*Offer1*), to their mutual contribution *Offer1*+*Offer2* which is essentially a standardized second stage gap whose values are in the range [-1,1]. We adopt this approach due to the symmetric property of our modified game, instead of using the ratio *Offer2*/(3 × *Offer1*) as the level of trustworthiness which is commonly adopted in the Trust game literature. Whenever *Offer2* = *Offer1* (including *Offer2* = *Offer1* = 0), the standardized gap equals zero, indicating that the responder is exactly reciprocal; and whenever *Offer2* > *Offer1* (*Offer2* < *Offer1*), the standardized gap is positive (negative), capturing the over- (under-) reciprocal behavior. The standardized gap takes the value of 1 only if player 2 received zero but altruistically sent back a positive amount, and takes the value of -1 only if player 2 received a positive amount but sent back nothing. This measurement thus reasonably helps address the possibility of receiving zero in the first stage and returning more than was received in the second stage.

Regardless of whether we use the absolute amount sent back (*Offer2*) after controlling for the amount received (*Offer1*) or use the standardized gap (*Offer2*- *Offer1*)/ (*Offer1*+ *Offer2*) to try to capture the possible gender difference within each treatment, the results remain qualitatively consistent with our main findings (see also Table A5 in Appendix).

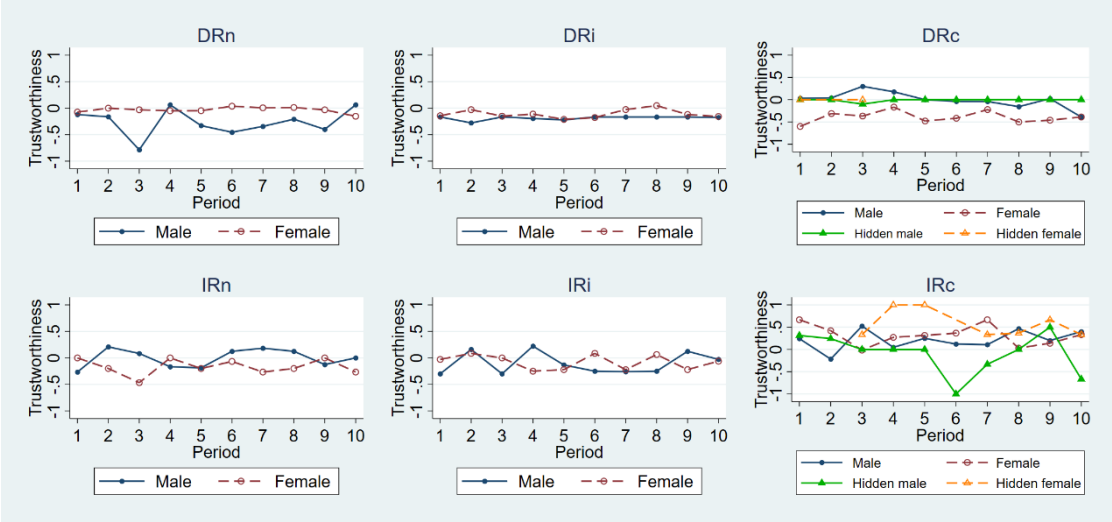
It is also noteworthy that although no substantial gender difference is detected in terms of trustworthiness in most of the treatments, men are robustly found to reciprocate slightly less than women in DRn, DRi and IRc, and slightly more than women in DRc and IRn, which corroborates the comparisons of unconditional second stage offers in Table 6. One possible explanation is that women may feel a greater obligation to return the favor in direct reciprocity treatments than in indirect reciprocity treatments (Buchan et al., 2008; Dufwenberg et al., 2001) as compared to men. The option of self-disclosure may alleviate their sense of obligation under direct reciprocity in which they are expected to pay back the sender's investment, while simultaneously promoting the obligatory feeling under indirect reciprocity, when they are in a chain to potentially socially transmit kind acts to others. Thus, it is possible that the process of voluntary information disclosure has the potential to reverse gender comparisons in reciprocal behavior.

Result 7 (Gender, Gender Information Disclosure, and Reciprocating): *Males and females are almost equally trustworthy, and self-disclosed responders are only marginally more reciprocal than the non-self-disclosed responders.*

The time paths of the trustworthiness measured by the standardized second stage gap, separated by gender, are shown in Figure 6. Overall, trustworthiness is smooth over periods especially for DRi, yet the variation in IRc is relatively large. The Wilcoxon rank-sum tests on the distribution of trustworthiness between periods 1–5 and periods 6–10 lend support to the idea that the reciprocal behavior is also fairly stable, regardless of gender and self-disclosure states (see Table A2 in Appendix).

Neither the interaction term between gender and period of play, nor between disclosure decision and period of plat demonstrated a significant coefficient once they were included in the regression model (not shown out of space consideration). Moreover, adding the lag amount returned (*Lag_offer2*) and the lag payoff (*Lag_payoff*) variables absorbs more variation (column (3) of Table 7), which again indicates that the behavior of responders is roughly consistent over time. A larger amount returned in the previous period is associated with a larger amount returned in the current period. A larger payoff in the previous period, that is, a larger lagged entrusted amount conditional on the lagged amount returned, tends to discourage player 2 from sending back more in the current period.

Figure 6. Trustworthiness over periods



3.2.1 Self-disclosure decisions of responders

Finally, we examine the self-disclosure decisions on the responder side during the pre-stage. Given our results so far, the self-disclosure rate is found to be relatively lower in IRc. A more detailed picture of the intertemporal development of subjects' self-disclosure decisions is shown in Figure 7. In the DRc treatment, we observe that in each period females almost always choose to disclose their gender, while males exhibit a relatively lower self-disclosure propensity, which is also stable over time. By contrast, subjects' self-disclosure decisions become more unpredictable when faced with reciprocating choices towards a third agent. The self-disclosure likelihood of women starts off exceeding that of men and exhibits substantial variation over the periods, which ends up being smaller

than that of men. It appears that the disclosure decisions of males are more discretionary, since they are generally more likely to resort to concealment, seemingly to help further anonymize the low offers they return. However, females' withholding their gender information does not harm their trustworthiness, which is consistent with prior observations that women are highly receptive to self-disclosure and adopt an affiliative style of self-presentation (Derlega et al., 1981). Yet their willingness to disclose information may decrease under indirect reciprocity due to the looser bond with their matched partners.

To understand the determinants of disclosure decisions, Table 8 provides estimations of Probit models with marginal effects reported and robust standard errors clustered at the subject level. Again, we find that the influence of responder gender is insignificant. However, a player who previously chose to conceal their gender is more likely to conceal gender again in the current period. In the DRc treatment, 5 out of the 13 responders chose to conceal their gender information at least once during the ten periods. However, this number doubles in the IRc treatment, and the average number of non-disclosure periods per subject increases as well. In addition, the significant coefficients for *Lag_offer1* and for the interaction term *Lag_offer1 ×Lag_disclosure* show that the probability of disclosure in the current period decreases with the lagged amount received if the responder's gender was concealed last period, while increasing with lagged amount received if the responder gender was revealed.

On closer inspection, when plotting the period-by-period adjustment of self-disclosure decisions in Figure 8, we observe that nearly half of subjects who concealed their gender in the previous period change to disclosure in the current period as long as the number of tokens received in the previous period is below 5. At the same time, around 10% of subjects switch to conceal their gender conditional on disclosure in the previous period. However, when the lagged amount received is at least as 5 tokens, the likelihood of reversing this decision is reduced to almost. In short, the median level of 5 tokens seems to serve as a natural reference point, below which subjects attempt to experiment with their disclosure decisions.

We also consider the set of independent variables which records the average amount received with (*Avgknown*) and without disclosing gender (*Avgunknown*) before that period, and the latest amount received with (*Latestknown*) and without (*Latestunknown*) disclosing gender before that period. Regressions (3) – (6) in Table 8 indicate that the self-disclosure likelihood is reinforced when rewarded with a larger number of tokens corresponding to previous disclosure, or punished with a lower number of tokens for non-disclosure, which is consistent with in the results of regression (2). The latest amount received without disclosing gender is more effective in changing the decisions of women compared with men.

Result 8 (Reciprocity Type, Time Dynamics, and Self-disclosure): *Self-disclosure propensity is lower and more volatile under indirect reciprocity than under direct reciprocity. Responders' decisions to disclose their gender is significantly associated with the amount they received in the past, mediated by their previous disclosure state.*

Figure 7. Self-disclosure probability over periods

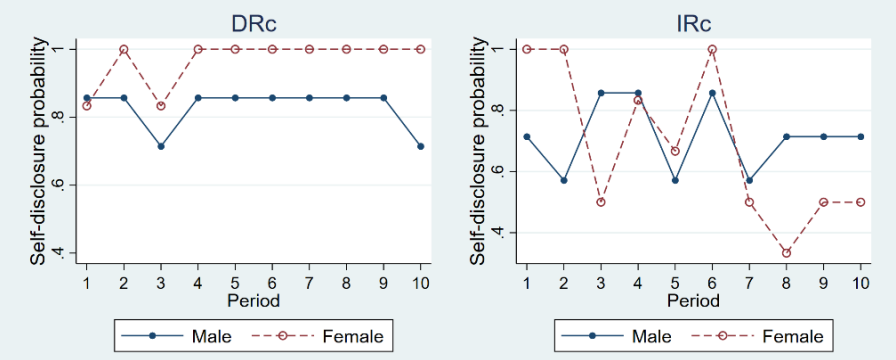


Figure 8. Adjustment to the self-disclosure decisions

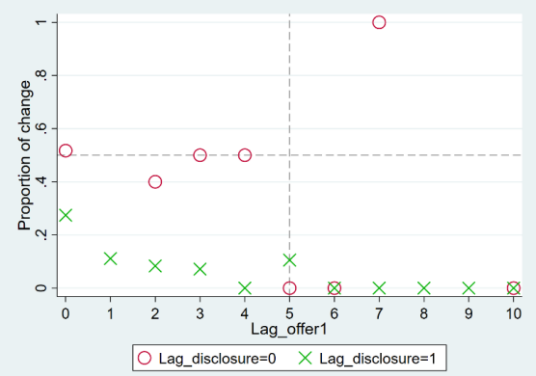


Table 8. Probit regression results for disclosure decisions

Disclosure	(1)	(2)	(3)	(4)	(5)	(6)
Period	-0.010	-0.008	-0.022	-0.022	-0.025*	-0.024*
	(0.009)	(0.008)	(0.014)	(0.014)	(0.013)	(0.013)
Indirect	-0.191**	-0.090	-0.252**	-0.301***	-0.067	-0.041
	(0.095)	(0.067)	(0.106)	(0.099)	(0.152)	(0.155)
Gender	-0.063	-0.017				
	(0.095)	(0.057)				
Lag_disclosure		0.110	-0.016	-0.017	0.126	0.112
		(0.070)	(0.085)	(0.073)	(0.084)	(0.088)
Lag_offer1		-0.033**				
		(0.016)				
Lag_offer1 ×Lag_disclosure		0.074***				
		(0.021)				
Avgknown			0.054*	0.036		
			(0.030)	(0.046)		
Avgunknown			-0.148***	-0.127***		
			(0.029)	(0.037)		
Gender ×Avgknown				0.025		
				(0.049)		
Gender ×Avgunknown				-0.047		

				(0.054)		
Latestknown					0.072***	0.109*
					(0.024)	(0.060)
Latestunknown					-0.068***	-0.393***
					(0.017)	(0.060)
Gender ×Latestknown						-0.047
						(0.068)
Gender ×Latestunknown						0.338***
						(0.072)
Observations	260	234	98	98	98	98
Pseudo R^2	0.070	0.284	0.432	0.438	0.325	0.337

Notes: Coefficients displayed are marginal effects. Standard errors clustered at the individual level are displayed in parentheses. *$p <$ 0.1, **$p < 0.05$, ***$p < 0.01$.

4. Conclusions and Discussion

As one of the potential methods to enhance gender equality of outcomes, gender-blind information policies and voluntary gender disclosure policies are intuitive propositions for the potential they hold to preclude explicit discrimination based on gender. However, the strategic considerations and dynamic consequences embedded in such policies have not been cleanly explored and tested. In a surplus generating experiment which incorporates both direct and indirect reciprocity, as well as gender information policies of 'no information', 'imposed information', and 'self-disclosed information', we study the consequences of such policies in a laboratory setting. To our knowledge, our study is unique to the literature in its endeavor to understand the effects of information policies in the gender context.

The experiments empirically show a number of regularities occurring with regard to gender differences in the two-stage trust and reciprocity game we implement. Firstly, male subjects in our study tended to give higher first-stage offers, apparently driven heavily by a direct reciprocity motive. Female subjects tend to give overall lower first-stage offers, although the discrepancy in given offers between direct reciprocity and indirect reciprocity environments is less than that of the male subjects. Overall, the imposed gender information condition yields the highest overall first stage surplus generation.

In the self-declared gender condition, both male and female subjects reacted negatively to partners who declined to reveal their gender, although male subjects were more prone to treating partners of unknown gender adversely specifically in the direct reciprocity setting, while female subjects tended to treat unknown gender partners similarly skeptically across the two reciprocity settings. Some of the reactions to subjects who decline to declare their gender may be due to the way in which subjects tended to use the gender information disclosure opportunity. Males are more likely than females to decline stating their gender throughout the ten periods of the direct reciprocity treatment, while female subjects nearly always disclose their gender, and converge quickly over the periods to unanimous disclosure. Quite a different pattern emerges in the indirect reciprocity treatment, in that female subjects begin the ten-period process almost unanimously disclosing their gender, and end the ten periods with a disclosure rate less than 50%. On their part, males maintain a slightly more stable but

lower disclosure rate than males' analogous behavior in the direct reciprocity environment. An examination of the dynamics of gender disclosure choices reveals that subjects tend to switch disclosure choices after receiving a below median transfer in the previous period, indicative of experimentation after a dissatisfactory result.

These differential treatments of individuals based on gender information in the first stage offers of the game, as well as the likely strategic motives behind disclosure choices in the self-declared condition, persist and are to some extent enhanced over the ten periods, even though receiving subjects are largely similar in reciprocity choices across gender and information. Altogether, what we observe is that although there are no direct payoff consequences to gender information, the effects of such information can alter the realized payoffs substantially as a result of subjects' *reactions* to the information or lack thereof. Combined with the higher non-disclosure behavior of males in the self-disclosure treatment, one interpretation is that perhaps knowing that female recipients tend to receive marginally more generous or equal first offer treatment by both genders, some male responders attempt to conceal their gender with the motive to create an ambiguity about their actual gender. However, the attempt is not effective on average, since senders of both genders tend to discriminate against such non-disclosure choices in the initial offers. Given that the second stage offers are rarely greater than first stage offers and nominal reciprocity is strong, the overall result of the gender disclosure choice setting is reduced surplus generated in the entire transaction compared to the setting of imposed information disclosure. Thus, our results contain a somewhat cautionary message with regard to voluntary disclosure policies in the gender domain; mainly that skepticism towards and exploitation of non-disclosure has the potential to reduce trust and social surplus.

Finally, we conclude with a brief comment about the external validity of our experiment to the context of different cultures and societies. Our experiments were implemented in China, which is currently a society characterized by relatively traditional gender roles compared to most industrialized nations. Although the labor force participation of women is high, the entry of women into significant leadership roles in different sectors of society and the economy is generally much rarer than in most Western countries, and may even be considered lower than in other industrialized Asian societies.[13] Thus, perhaps a straightforward hypothesis about the external validity of our study is to those societies which are similar to China in terms of having relatively traditional gender notions, while also having approximate gender parity in other basic domains such as labor force participation and educational attainment. The results obtained here could also be predictive for societies which are traditional in gender attitudes but are moving towards gender equality in education and the labor market, similar to in the Chinese context. Further work in this direction testing the corresponding effects under different social gender norms would be valuable to implement in varying cultural settings.

[13] A relevant experimental study on gender differences in leadership framing is Jiang and Wang (2020).

References

1. Andreoni, J., Vesterlund, L., 2001. Which is the fair sex? Gender differences in altruism. *Quarterly Journal of Economics* 116, 293–312.
2. Bellemare, C., Kröger, S., 2007. On representative social capital. *European Economic Review* 51(1), 183–202.
3. Ben-Ner, A., Putterman, L., Kong, F., Magan, D., 2004. Reciprocity in a two-part dictator game. *Journal of Economic Behavior and Organization* 53 (3), 333–352.
4. Berg, J., Dickhaut, J. W., McCabe, K. A., 1995. Trust, reciprocity, and social history. *Games and Economic Behavior* 10 (1), 122–142.
5. Bohnet, I., Hermann, B., Zeckhauser. R., 2010. The requirements for trust in Gulf and Western countries. *Quarterly Journal of Economics* 125 (2), 811–828.
6. Buchan, N., Croson, R., Dawes, R. M., 2002. Swift neighbors and persistent strangers: A cross-cultural investigation of trust and reciprocity in social exchange. *American Journal of Sociology* 108 (1), 168–206.
7. Buchan, N. R., Croson, R., Solnick, S. J., 2008. Trust and gender: An examination of behavior and beliefs in the investment game. *Journal of Economic Behavior and Organization* 68 (3–4), 466–476.
8. Chaudhuri, A., Gangadharan, L., 2007. An experimental analysis of trust and trustworthiness. *Southern Economic Journal* 73 (4), 959–985.
9. Chaudhuri, A., Sbai, E. 2011. Gender differences in trust and reciprocity in repeated gift exchange games. New Zealand Economic Papers: Special Issue in *Economic Psychology and Behavioral Economics* 45(1–2), 81–95.
10. Chaudhuri, A., Paichayontvijit, T., Shen, L., 2013. Gender differences in trust and trustworthiness: Individuals, single sex and mixed sex groups. *Journal of Economic Psychology* 34, 181–194.
11. Cox, J. C., Deck C. A., 2006. When are women more generous than men? *Economic Inquiry* 44 (4), 587–598.
12. Croson, R., Buchan, N., 1999. Gender and culture: International experimental evidence from trust games. *American Economic Review* 89 (2), 386–391.
13. Croson, R., Gneezy, U., 2009. Gender Differences in Preferences. *Journal of Economic Literature* 47 (2), 1–27.
14. Derlega, V. J., Durham, B., Goekel, B., Sholis, D., 1981. Sex differences in self-disclosure: Effects of topic content, friendship, and partner's sex. *Sex Roles* 7 (4), 433–447.
15. Dittrich, M., 2015. Gender differences in trust and reciprocity: evidence from a large-scale experiment with heterogeneous subjects. *Applied Economics* 47 (34–36), 3825–3838.
16. Drouvelis, M., Gerson, J., Powdthavee, N., Riyanto, Y.E., 2020. Large Losses from Little Lies: Randomly Assigned Opportunity to Misrepresent Substantially Lowers Later Cooperation and Worsens Income Inequality. *IZA Discussion Papers*, No. 13619, Institute of Labor Economics (IZA), Bonn.
17. Dufwenberg, M., Gneezy, U., Güth, W., van Damme, E., 2001. Direct versus indirect reciprocity: An experiment. *Homo Oeconomicus* 18, 19–30.
18. Dufwenberg, M., Muren, A., 2006. Generosity, anonymity, gender. *Journal of Economic Behavior and Organization* 61(1), 42–49.
19. Eckel, C. C., Grossman, P. J., 2001. Chivalry and Solidarity in Ultimatum Games. *Economic Inquiry* 39 (2), 171–188.
20. Eckel, C. C., Grossman, P. J., 2008. Differences in the economic decisions of men and women: Experimental evidence. In C. Plott & V. Smith (Eds.). Handbook of experimental economics results (Vol. 1, pp. 509–519). Amsterdam: North-Holland.
21. Eckel, C. C., Wilson, R. K., 2005. Attractiveness and trust: Does beauty confound intuition? Virginia Tech University Department of Economics Working Paper.
22. Eckel, C. C., Wilson, R. K., 2003. Conditional trust: Sex, race and facial expressions in an experimental trust game. Virginia Tech University Department of Economics Working Paper.

23. Engelmann, D, Fischbacher, U., 2009. Indirect reciprocity and strategic reputation building in an experimental helping game. *Games and Economic Behavior* 67 (2), 399–407.
24. Fischbacher, U., 2007. Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10 (2), 171–178.
25. Garbarino, E., Slonim, R., 2009. The robustness of trust and reciprocity across a heterogeneous U.S. population. *Journal of Economic Behavior and Organization* 69 (3), 226–240.
26. Gilligan, C., 1982. In a different voice: Psychological theory and women's development. Cambridge and London: Harvard University Press.
27. Goldin, C., Rouse, C., 2020. Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians, American Economic Review, 90 (4), 715 – 741.
28. Greiner, B., Levati, M.V., 2005. Indirect reciprocity in cyclical networks: An experimental study. *Journal of Economic Psychology* 26, 711 – 731.
29. Herne, K., Lappalainen, O., Kestilä-Kekkonen, E., 2013. Experimental comparison of direct, indirect and serial reciprocity. *Journal of Socio-Economics* 45, 38–46.
30. Ingram, B. L., Berger, S. E., 1977. Sex-role orientation, defensiveness, and competitiveness in women. *Journal of Conflict Resolution* 21 (3), 501–518.
31. Innocenti, A., Pazienza, M. G., 2006. Altruism and gender in the trust game. University of Sienna Department of Economics Working Paper 5/2006.
32. Jiang, X., Wang, X., Gender Difference in Leaders Framed with Responsibility or Authority: An Experimental Study. Working Paper, 2020.
33. Kamei, K., 2020. Voluntary disclosure of information and cooperation in simultaneous-move economic interactions. *Journal of Economic Behavior and Organization* 171, 234–246.
34. Li, J., 2018. Direct reciprocity and serial reciprocity: an experimental analysis. *Singapore Economic Review*, 1–20.
35. McAllister, H. A., 1980. Self-disclosure and liking: Effects for senders and responders. *Journal of Personality* 48, 409–418.
36. McGee, P., Constantinides, S., 2013. Repeated play and gender in the ultimatum game. *The Journal of Socio-Economics* 42, 121–126.
37. Migheli, M., 2007. Trust, gender, and social capital: Experimental evidence from three western European countries. Unpublished.
38. Miller, L., Ubeda, P., 2012. Are women more sensitive to the decision-making context? *Journal of Economic Behavior and Organization* 83 (1), 98–104.
39. Mujcic, R., Leibbrandt, A., 2017. Indirect reciprocity and prosocial behavior: Evidence from a natural field experiment. *The Economic Journal* 128, 1683–1699.
40. Okada, I., 2020. A Review of Theoretical Studies on Indirect Reciprocity. *Games*, 11(27).
41. Riyanto, Y. E., Zhang, J., 2014. An egalitarian system breeds generosity: The impact of redistribution procedures on pro-social behavior. *Economic Inquiry* 52 (3), 1027–1039.
42. Schwieren, C., Sutter, M., 2008. Trust in cooperation or ability? An experimental study on gender differences. *Economics Letters* 99, 494–497.
43. Servátka, M. 2009. Separating reputation, social influence, and identification effects in a dictator game. *European Economic Review* 53 (2), 197–209.
44. Small, D.A., Loewenstein, G., 2003. Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and Uncertainty* 26 (1), 5–16.
45. Snijders, C., Keren, G., 2001. Do You Trust? Whom Do You Trust? When Do You Trust? *Advances in Group*

Processes 18, 129–160.

46. Solnick, S. J., 2001. Gender differences in the ultimatum game. *Economic Inquiry* 39 (2), 189–200.
47. Stanca, L., 2009. Measuring indirect reciprocity: whose back do we scratch? *Journal of Economic Psychology* 30 (2), 190–202.
48. Stanca, L., Bruni, L., Corazzini, L., 2009. Testing theories of reciprocity: do motivations matter? *Journal of Economic Behavior and Organization* 71, 233–245.
49. Van Apeldoorn, J., Schram, A., 2016. Indirect Reciprocity: A Field Experiment, *PLOS One* 11(4): e0152076.
50. Xie, Z., Page, L. and Hardy, B., 2017. Investigating Gender Difference Under Time Pressure in Financial Risk Taking. *Frontiers of Behavioral Neuroscience* 11, Article 246.
51. Zeltzer, D., 2020. Gender Homophily in Referral Networks: Consequences for the Medicare Physician Earnings Gap. *American Economic Journal: Applied Economics* 12 (2), 169 – 197.

Appendix A. Additional results

Table A1. Wilcoxon rank-sum tests on first stage offers by phase

	DRn	*DRi*	*DRc*	*IRn*	*IRi*	*IRc*
Aggregate	p = 0.039	p = 0.673	p = 0.994	p = 0.395	p = 0.789	p = 0.129
Male	p = 0.023	p = 0.393	p = 0.974	p = 0.547	p = 0.605	p = 0.406
Male-to-male		p = 0.530	p = 0.337		p = 0.995	p = 0.297
Male-to-female		p = 0.417	p = 0.800		p = 0.705	p = 0.526
Male-to-unknown			p = 0.731			p = 0.297
Female	p = 0.679	p = 0.244	p = 0.967	p = 0.563	p = 0.956	p = 0.246
Female-to-male		p = 0.238	p = 0.995		p = 0.247	p = 0.169
Female-to-female		p = 0.268	p = 0.711		p = 0.656	p = 0.407
Female-to-unknown			p = 0.541			p = 0.549

Note: The Wilcoxon rank-sum tests for pairwise comparisons on the first stage offer broken up by two phases (periods 1-5 vs periods 6-10) are reported in the table.

Table A2. Wilcoxon rank-sum tests on second stage offers by phase

	DRn	*DRi*	*DRc*	*IRn*	*IRi*	*IRc*
Aggregate	p = 0.893	p = 0.7061	p = 0.277	p = 0.426	p = 0.610	p = 0.655
Male (Unhidden)	p = 0.966	p = 0.596	p = 0.101	p = 0.286	p = 0.482	p = 0.748
Hidden male			p = 0.317			p = 0.140
Female (Unhidden)	p = 0.813	p = 0.961	p = 0.919	p = 0.830	p = 0.810	p = 0.956
Hidden female			-			p = 0.329

Note: The Wilcoxon rank-sum tests for pairwise comparisons on the second stage offer broken up by two phases (periods 1-5 vs periods 6-10) are reported in the table.

Table A3. Tobit regression results for first stage offer adjustments (DRi & IRi)

	Model I				Model II			
Dep. Var	Offer11		Offer10		Offer11		Offer10	
Period	-0.091	-0.101	-0.033	-0.053	0.021	0.023	-0.007	-0.002
	(0.133)	(0.134)	(0.068)	(0.068)	(0.097)	(0.099)	(0.076)	(0.071)
Indirect	-1.501	-1.942*	-1.705*	-2.331**	-1.232	-1.416	-1.586*	-1.937*
	(0.939)	(1.063)	(0.962)	(1.094)	(0.880)	(0.906)	(0.945)	(1.057)
Lag_offer11	0.860***	0.778***			0.943***	0.915***		
	(0.131)	(0.151)			(0.152)	(0.157)		
Lag_offer10			0.980***	0.796***			1.076***	1.000***
			(0.151)	(0.152)			(0.143)	(0.152)
Avgmen	0.613***	0.662***	0.216**	0.473***				
	(0.220)	(0.238)	(0.108)	(0.156)				
Avgwomen	0.030	-0.197	0.189	-0.173				
	(0.165)	(0.193)	(0.168)	(0.164)				
Gender ×Avgmen		0.042		-0.235*				
		(0.177)		(0.120)				
Gender ×Avgwomen		0.258		0.589***				
		(0.228)		(0.187)				
Latestmen					0.470***	0.414***	0.115	0.119
					(0.175)	(0.153)	(0.085)	(0.130)
Latestwomen					0.103	0.075	0.164	0.042
					(0.099)	(0.131)	(0.107)	(0.145)
Gender ×Latestmen						0.103		0.035
						(0.161)		(0.127)
Gender ×Latestwomen						0.025		0.194
						(0.161)		(0.161)
Constant	-1.015	-0.391	-0.708	0.103	-2.078	-1.848	-0.877	-0.471
	(1.539)	(1.592)	(1.189)	(1.345)	(1.364)	(1.366)	(1.090)	(1.216)
Observations	174	174	174	174	174	174	174	174
Left-censored	53	53	50	50	53	53	50	50
Right- censored	42	42	43	43	42	42	43	43
Pseudo R^2	0.290	0.298	0.329	0.349	0.299	0.301	0.328	0.335

Notes: *Offer11* denotes the first stage offer to male responders, and *Offer10* denotes the first stage offer to female responders. Standard errors clustered at the individual level are displayed in parentheses. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table A4. Tobit regression results for first stage offer adjustments (DRc & IRc)

	Model I			Model II		
Dep. Var	Offer12	Offer11	Offer10	Offer12	Offer11	Offer10
Period	0.005	-0.418**	-0.191	-0.020	-0.392**	-0.178
	(0.154)	(0.185)	(0.164)	(0.122)	(0.187)	(0.157)
Indirect	-1.003	-1.717	-1.458**	-1.005	-1.419*	-1.090*
	(1.027)	(1.043)	(0.727)	(1.012)	(0.843)	(0.624)
Lag_offer12	1.264***			1.263***		
	(0.255)			(0.214)		
Lag_offer11		1.290***			1.243***	
		(0.245)			(0.188)	
Lag_offer10			1.157***			1.187***
			(0.137)			(0.131)
Avgunknown	0.175	0.160	0.090			
	(0.169)	(0.183)	(0.132)			
Avgmen	0.044	-0.305	-0.261			
	(0.147)	(0.237)	(0.193)			
Avgwomen	0.014	0.283	0.302*			
	(0.115)	(0.192)	(0.175)			
Latestunknown				0.108	0.151	0.119
				(0.110)	(0.160)	(0.134)
Latestmen				0.024	-0.126	-0.108
				(0.085)	(0.110)	(0.116)
Latestwomen				0.055	0.186**	0.144
				(0.086)	(0.094)	(0.095)
Constant	-2.596	1.773	0.796	-2.145	1.362	0.316
	(1.785)	(1.545)	(1.360)	(1.482)	(1.065)	(1.005)
Observations	109	109	109	112	112	112
Left-censored	62	51	39	62	51	39
Right- censored	11	16	19	11	16	19
Pseudo R^2	0.403	0.376	0.363	0.396	0.374	0.349

Notes: *Offer12* denotes the first stage offer to non-disclosed responders, *Offer11* denotes the first stage offer to known male responders, and *Offer10* denotes the first stage offer to known female responders. Standard errors clustered at the individual level are displayed in parentheses. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table A5. Tobit regression results for reciprocation by treatment

Second stage offer	(1)	(2)	(3)	(4)	(5)	(6)
	DRn	DRi	DRc	IRn	IRi	IRc
Offer1	1.188***	1.313***	1.473***	1.132***	0.390	0.780***
	(0.209)	(0.202)	(0.155)	(0.264)	(0.236)	(0.271)
Period	0.030	-0.169	-0.185	-0.298	-0.251	-0.170
	(0.110)	(0.176)	(0.161)	(0.272)	(0.231)	(0.165)
Gender	-2.600*	-0.971	3.852	2.538	2.604	-2.470
	(1.540)	(2.408)	(2.449)	(2.193)	(4.331)	(3.277)
Constant	-0.200	-1.271	-4.762**	-3.909**	-1.901	2.823
	(1.209)	(1.132)	(2.321)	(1.868)	(2.912)	(2.172)
Left-censored	15	26	44	87	73	45
Right- censored	43	41	35	6	9	22
Pseudo R^2	0.144	0.126	0.240	0.103	0.017	0.033
Standardized gap	(1)	(2)	(3)	(4)	(5)	(6)
	DRn	DRi	DRc	IRn	IRi	IRc
Period	0.002	0.005	-0.023**	0.008	0.001	-0.014
	(0.006)	(0.009)	(0.011)	(0.013)	(0.025)	(0.019)
Gender	-0.249***	-0.094	0.428**	0.198	-0.036	-0.466
	(0.094)	(0.191)	(0.196)	(0.139)	(0.386)	(0.333)
Constant	-0.045	-0.145**	-0.308*	-0.244**	-0.122	0.759***
	(0.094)	(0.064)	(0.172)	(0.111)	(0.304)	(0.265)
Left-censored	8	15	21	16	28	9
Right- censored	1	0	2	9	18	43
Pseudo R^2	0.095	0.013	0.145	0.017	0.0001	0.023
Observations	130	130	130	130	130	130

Notes: Standard errors clustered at the individual level are displayed in parentheses. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Appendix B. Experimental instructions

Thank you for your participation in this experiment! Please read the following instructions carefully. If you have any questions, please feel free to ask us. Please note that you cannot communicate with other participants during the experiment.

You will be paid to complete the experiment according to the instructions. Your payoff for the experiment will be determined by your choices and the choices of other participants. At the end of the experiment the tokens that you have earned will be converted into CNY at the exchange rate *1 token = 1 CNY*.

Matching rules [common to all the treatments]

There are *10 periods of repeated games* in total. At the beginning of the experiment, you will be randomly assigned a role (player 1 / player 2) with equal probability. You will form a group with one of the participants *in another different role* to carry out the experimental task anonymously.

Your role *remains the same* for the 10 periods of play. At the beginning of each period, the composition of each group will be *randomly re-formed*. That is, your paired member will change in different periods, and the formation of pairs in different periods is completely independent of each other.

Playing rules [specific to treatments DRc, IRc]

The experiment will take place in three stages chronologically, which are Player 2 Disclosure Stage, Player 1 Allocation Stage and Player 2 Allocation Stage. The details are as follows:

0. Pre-stage: Player 2 Disclosure Stage

At this stage, the gender of player 2 will be accessed by the program, and player 2 will have to decide whether to disclose his/her gender to the paired member in the upcoming period of play.

1. First stage: Player 1 Allocation Stage

At the beginning of this stage, player 1 will receive an endowment of *10 tokens*, and player 2 will receive *zero*. Then player 1 will have to decide independently how many tokens (X) to send to player 2, and the remaining tokens (10-X) will be deposited in his/her own account. Player 2 will receive *3 tokens* for each token sent by player 1. X should be an integer between [0,10].

Player 1 will have to make allocation decisions according to three possible situations:

- Player 2 does not disclose his/her gender;
- Player 2 discloses his gender and is male;
- Player 2 discloses her gender and is female.

The actual number of tokens (X) sent to player 2 will be determined by the player 2's disclosure

decision in the pre-stage and gender.

Player 2 will wait for player 1 to make an allocation decision at this stage. After player 1 completes the decision, player 2 will be informed of the number of tokens allocated by player 1 (X) and the account balance in the first stage (3X); Player 1 will be informed whether player 2 chooses to disclose his/her gender, his/her real gender conditional on disclosure, the number of tokens finally allocated to player 2 (X), and the account balance in the first stage (10-X).

2a. Second stage: Player 2 Allocation Stage [specific to treatments DRn, DRi, DRc]

At the beginning of this stage, player 2 will receive an endowment of *10 tokens*, and player 1 will receive *zero*. Then player 2 will have to decide independently how many tokens (Y) to send to player 1, and the remaining tokens (10-Y) will be deposited in his/her own account. Player 1 will receive *3 tokens* for each token sent by player 2. Y should be an integer between [0,10].

Player 1 will wait for player 2 to make an allocation decision at this stage. After player 2 completes the decision, player 1 will be informed of the number of tokens allocated by player 2 (Y) and the account balance in the second stage (3Y); Player 2 will be informed of the number of tokens allocated to player 1 (Y), and the account balance in the second stage (10-Y).

The payoff you get for each period of play equals the sum of the account balance in the two stages:

$$Payoff\ in\ each\ period\ = \begin{cases} (10 - X) + 3Y, & \text{if you are player 1} \\ 3X + (10 - Y), & \text{if you are player 2} \end{cases}$$

2b. Second stage: Player 2 Allocation Stage [specific to treatments IRn, IRi, IRc]

Before the start of this stage, all pairs will be *randomly and independently re-matched* again. This means your partner in the second stage will be *different* from that in the first stage. Player 1 will be informed of the disclosure decision made by the *randomly re-matched* player 2 during the pre-stage and his/her real gender conditional on disclosure.

At the beginning of this stage, player 2 will receive an endowment of *10 tokens*, and player 1 will receive *zero*. Then player 2 will have to decide independently how many tokens (Y) to send to a *randomly re-matched* player 1', and the remaining tokens (10-Y) will be deposited in his/her own account. Player 1' will receive *3 tokens* for each token sent by player 2. Y should be an integer between [0,10].

Player 1 will wait for player 2 to make an allocation decision at this stage. After player 2 completes the decision, player 1 will be informed of the number of tokens allocated by a randomly re-matched player 2' (Y') and the account balance in the second stage (3Y'); Player 2 will be informed of the number of tokens allocated to a randomly re-matched player 1' (Y), and the account balance in the second stage (10-Y).

The payoff you get for each period of play equals the sum of the account balance in the two stages:

$$Payoff\ in\ each\ period\ = \begin{cases} (10 - X) + 3Y', & \text{if you are player 1} \\ 3X + (10 - Y), & \text{if you are player 2} \end{cases}$$

Payment rules [common to all the treatments]

We will randomly select *one period* from the 10 periods of play, and take your payoff in that period plus a 10 CNY show-up fee as your final earnings in the experiment. The tokens that you have earned will be converted into CNY at the exchange rate *1 token = 1 CNY*. The resulting amount will be paid to you via *Wechat* at the end of the experiment.

If you have any questions, please raise your hand to the experimenter. If there is no problem, we will start the experiment after all participants confirm.

[Notes. The instructions for the other treatments are almost identical, with the following slight changes:

- In the part of *Matching rules*: subjects in the indirect reciprocity treatments (IRn, IRi, IRc) are additionally told that their paired members will also change before the start of the second stage within each period, and the formation of pairs in the second stage is completely independent of that in the first stage.
- In the part of *Playing rules*: There is no *Pre-stage* in the no information and imposed information treatments.
- In the *First stage* of no information treatments (DRn & IRn), player 1 will have to directly decide on the number of tokens (X) sent to player 2, and only be informed of the number of tokens allocated to player 2 (X) as well as the account balance in the first stage (10-X) later on.
- In the *First stage* of imposed information treatments (DRi & IRi), player 1 will have to make allocation decisions according to two possible situations:
  - Player 2 is male;
  - Player 2 is female.

  The actual number of tokens (X) sent to player 2 will be determined by the gender of player 2. After player 1 completes the decision, he will be informed of the gender of player 2, the number of tokens finally allocated to player 2 (X), and the account balance in the first stage (10-X).
- 2a and 2b are instructions for the *Second stage* of the direct reciprocity treatments and indirect reciprocity treatments respectively. **]**