Gender Differences in Reactions to Enforcement Mechanisms: A Large-Scale Natural Field Experiment

Zhengyang Bao and Difang Huang

June 18, 2023

Abstract

We randomize dunning text messages to 17,545 borrowers from a FinTech lending platform to study gender differences in responsiveness to enforcement mechanisms. Compared to a reminder message, messages inducing social pressure and financial incentives can further reduce the overdue rate. However, women are more responsive to social pressure, while men are more sensitive to financial incentives. These results hold for borrowers with different credit risks and are robust to all observable control variables and matching methods. Endorser choices and financial incentive sizes cannot fully resolve above gender differences. These findings suggest some seemingly gender-neutral practices may create unintended gender disparities.

JEL Code: C93, D91, J16.

Keywords: Gender Differences; Natural Field Experiment; FinTech.

I. Introduction

Men and women have some fundamental differences in preferences and behaviors. For example, women tend to be more risk-averse, less likely to enter competitive environments, and respond to social and financial incentives differently from men (e.g., Andreoni and Vesterlund, 2001; Gneezy and Rustichini, 2004). These gender differences hint that men and women may react to some enforcement mechanisms differently. While the understanding of such gender differences may help policymakers to improve the implementation of laws and regulations, empirical evidence on this is scarce. We fill this gap by conducting a natural field experiment (Harrison and List, 2004) investigating potential gender differences in response to enforcement mechanisms that are commonly employed in financial contexts. In particular, we consider enforcement mechanisms based on social and financial incentives and test them on borrowers from a large FinTech lending platform in China.[1]

The FinTech lending platform collects funds from lenders and provides a loan facility to borrowers, similar to a financial intermediary. It attracted 6.54 million users from all over China and had a transaction volume of RMB 17.6 billion (USD 2.64 billion) in the year 2017 when the experiment was conducted. The platform is an ideal test bed for investigating gender differences in response to enforcement mechanisms, for several important reasons. First, the gender information is highly reliable, as each borrower must provide the platform with her/his unique national ID, which contains information about the person's gender at birth. Second, the non-compliance (overdue) rate is relatively gender-neutral prior to the experiment (according to all credit records in 2016, male borrowers account for about 55% of the late repayments), so the results are unlikely to suffer from biases caused by selection or stereotypes. Third, the platform provides a large sample size with a moderate overdue

[1]The popularity of FinTech lending is not restricted to China. de Roure et al. (2021) provide empirical evidence that FinTech lending in Germany substitutes for traditional banking by providing high-risk borrowers access to credit. Di Maggio and Yao (2021) show that in the US, FinTech lending enables borrowers with immediate consumption needs to finance their expenses beyond capability.

rate, so we have a reasonable number of observations to make statistical inferences. Fourth, overdue behavior is easily identifiable, minimizing the bias caused by measurement error. Fifth, the project may find a practical gender-specific mechanism to deter late repayments, which would be an economically and socially meaningful outcome.

In the experiment, male and female borrowers are randomized into one of the treatment groups and receive text messages from the platform asking them to repay on time. The content of the messages varies across treatments: we embed into each message one of five enforcement mechanisms that are commonly employed in other contexts and likely to create gender-specific reactions.[2] The first mechanism is a simple reminder message. The second and third mechanisms invoke social pressure. Specifically, we test how borrowers react to information about a descriptive social norm (i.e., most borrowers repay on time) and a threat to inform the endorsers (i.e., friends and family members the borrower nominates when initiating the loan) if they fail to repay on time. The fourth and fifth mechanisms introduce financial incentives. We test rewards for compliance (i.e., the platform offers to decrease the interest rate for future loans by 5% if the borrower makes the repayment on time) and punishments for non-compliance (i.e., the platform threatens to increase the interest rate for future loans by 5% if the borrower fails to make the repayment on time).

We find that social pressure and financial incentives successfully reduce the overdue rate for both male and female borrowers compared with the scenarios where no message is sent or only a reminder message is sent. However, there are significant gender differences in the responsiveness to these incentives. The most effective mechanism to encourage female borrowers to repay loans on time is the social pressure from notifying the endorsers, while the most effective mechanism for their male counterparts is the threat of financial punishment. If we send the same message to both male and female borrowers, we can decrease the overdue rate to 17.0% at best. However, if we tailor the message according to the borrower's gender,

[2]Hypotheses based on pre-existing gender differences and literature predict gender-specific responsiveness to these enforcement mechanisms. Details on these hypotheses are described in Section II.D.

we can further cut the overdue rate to 13.7%.

One potential concern about the analyses above is that pre-existing heterogeneity between male and female borrowers might affect our findings. We address it by applying the propensity score matching and the entropy balancing methods based on the full set of observed demographic and credit characteristics. The results based on the matched samples confirm our main findings. The effectiveness of some of the enforcement mechanisms is not gender-neutral; applying a uniform mechanism may create unintended gender gaps, while a gender-dependent enforcement mechanism may have greater merit in deterring non-compliance.

After observing gender differences at the average level, a natural follow-up question is whether a subgroup of borrowers drives the result. In particular, borrowers at the margin of delinquent are more likely to be affected by the message. To tackle this question, we further study whether borrowers' credit risk interacts with gender differences in responsiveness to text messages. By regressing the overdue indicator on a complete set of observed demographic and credit characteristics of borrowers from the group which did not receive any text message from the platform, we estimate the credit risk for each borrower. The data shows that the gender differences apply to both high- and low-credit risk borrowers.

Conducting further analyses, we assess potential explanations for these findings. First, we explore whether the choice of endorsers is related to the gender gap in the social incentive treatments. The data observe that the non-compliance rate is negatively correlated with the number of family endorsers, as opposed to friend endorsers, and that the message warning that the platform may notify the endorsers about the potential late payment strengthens the negative relationship between the number of family endorsers and the overdue probability. This explains part of the gender gap in the treatment sending shaming messages, as female borrowers nominate significantly more family members as endorsers. Second, we investigate whether borrowers respond to the (expected) size of financial incentives. As the monetary value of the financial incentive is proportional to the interest rate and expected future borrowing amount by design, we use the interest rate and proxies for future borrowing amount to capture the value of financial incentives. The results show that borrowers only respond to the size of financial incentives measured by interest rate but not other measures. Therefore, the incentive size channel may not explain the gender gaps in the reward and punishment treatments. Rather, the data indicate that men are more responsive to the existence of financial incentives per se rather than the size of it, which explains part of the overall gender gap in the treatment effects.

This project extends the growing literature studying gender differences, and in particular how men and women differ in preferences and behaviors (Bao and Huang, 2022; Barber et al., 2021; Bursztyn et al., 2017, 2020; Gneezy and Rustichini, 2004; Niederle and Vesterlund, 2007).[3] We complement this literature by finding gender differences in reactions to enforcement messages as women are more responsive to social pressure, while men are more sensitive to financial incentives. We also uncover new channels, through the choice of endorsers and responsiveness to the existence of incentives, that contribute to the gender difference in the outcome. Moreover, this paper adds to the ongoing equality-equity discussion in a financial environment.

In addition, we supplement the nascent literature on financial institution practices and borrower behaviors (Bao and Huang, 2021; Bursztyn et al., 2018, 2019b; Di Maggio and Yao, 2021). In particular, Bursztyn et al. (2019b) conduct an experiment to investigate the effectiveness of using a moral appeal in reducing the delinquency rate among credit card customers at a large Islamic bank in Indonesia. Bao and Huang (2021) compare bank and FinTech lending during COVID-19 and find that FinTech lenders provide more credit access after the pandemic while being more susceptible to the increased credit risk. Observable heterogeneity in borrower and loan characteristics cannot fully explain the differences; borrowers holding both loan types prioritize bank loan payments. We add to the study of borrowing behavior in FinTech lending by discovering that both FinTech dunning messages

[3]Croson and Gneezy (2009) provide a comprehensive review of gender differences in preferences.

and borrowers' genders can affect the delinquency ratio.

This project also contributes to the literature on using messages to influence behaviors. For instance, messages that provoke pro-social preferences such as morality, peer effects, and social image concerns have been widely applied to achieve desirable social and economic outcomes (Allcott, 2011; Karlan et al., 2012).[4] We, therefore, extend this line of literature by testing additional incentives, considering a novel gender angle, checking under-examined channels that drive gender differences, and exploring an emerging financial market. More broadly, this paper is linked to the literature on the design of deterrence mechanisms using financial and social incentives (Abbink et al., 2014; Bursztyn et al., 2019a; Xiao and Houser, 2011). We compare the efficacy of several approaches and discuss how to utilize gender information to improve the deterrence effect. Our paper also improves the understanding of underlying factors affecting the performance of some enforcement mechanisms.

Last but not least, we supplement the empirical literature studying the effectiveness of peer effects (Bursztyn and Jensen, 2015; DellaVigna et al., 2012). For example, Bursztyn et al. (2014) find that herding in asset markets can be partially explained by social learning and social utility. We add to this literature by comparing the effectiveness of messages revealing peer borrowers' repayment behavior with other types of messages and focusing on a novel gender angle.

The remainder of the paper is organized as follows. Section II describes the design of the experiment and the development of hypotheses. Section III discusses our main experimental results and potential explanations for these findings. Finally, Section IV provides concluding remarks.

[4]We refer interested readers to a review of the literature by Bursztyn and Jensen (2017).

II. Experimental Design

A. The FinTech lending platform

In China, many individuals have limited access to the formal banking system. FinTech lending has evolved to meet the needs of people who cannot access enough credit from banks. Compared with traditional banks, FinTech lending companies operate digitally, process loan applications faster, and provide more flexible and personalized options to borrowers (such as loan conditions and collateral requirements). The FinTech lending platform that provides the case setting for this study matches borrowers with lenders nationwide and charges fees for the service. Figure 1 plots the geographic distribution of male and female borrowers in our sample.

[Place Figure 1 about here]

The platform facilitates several types of loans, including collateral-based loans and unsecured loans. There are also options for different amortization schedules. We experiment on the borrowers who take out credit loans with the principal and interest paid at maturity because this loan type is the most straightforward loan arrangement, and incentives are less likely to be confounded with other factors.

A.1. The borrowing process

Before taking out a loan, a borrower has to register with the platform by providing personal information, including a copy of the national ID, which contains demographic information, including gender, age, and birthplace. The borrower must also nominate five friends and/or family members as endorsers and submits their contact information to the platform. The platform then contacts all proposed endorsers to verify that they are genuinely related to the borrower. The platform may contact the endorsers if the borrower fails to pay the money back, but they are not liable for the unpaid debts. A detailed description of variables collected by the platform appears in Appendix A.1.

After completing the registration process, a prospective borrower has to submit an application specifying the amount of money s/he would like to borrow and the loan duration (term). Borrowers can choose a loan amount ranging from RMB 100 to RMB 500,000, but more than 90% of them choose an amount below RMB 10,000. Borrowers can also choose a loan duration of one month to twenty-four months. The platform evaluates the application and chooses to either reject the application or accept the application and set an interest rate for the prospective loan. In a typical situation, the interest rate of the loan is the same as its' immediate predecessor. The platform negotiates the interest rate with a first-time borrower and may increase the rate for a borrower whose previous loan is overdue. The annualized interest rate usually ranges from 7% to 24%. The platform then posts the detailed borrowing request online, where potential lenders can view the application and decide whether to lend money to the borrower. The borrower gets the loan if the lender(s) decide to fulfill the requested amount. The interest is accumulated on a daily basis, and the borrower must pay back the principal and interest in full to the platform before the term ends. They cannot pay back in stages as no partial payments are allowed. Each borrower can have only one outstanding loan at any one time; this prevents the borrower from seeking a new loan from the platform to pay back the existing one.

If a loan is overdue, the borrower is given a 29-day window to repay the money with a daily penalty that is much higher than the interest rate. If the borrower still fails to pay after the 29-day window, the loan is considered a default. The platform may pursue the repayment through legal procedures.

B. The treatments

Limited attention and moral hazard are two likely contributors to late repayments of loans. To mitigate forgetfulness, we design a reminder treatment (hereafter Reminder), which sends a reminder text to the borrower one day before the repayment is due. There are several ways to reduce moral hazard. We use social pressure and monetary incentives in the experiment. To induce social pressure in the social norm treatment (hereafter Norm), we send the borrower a message stating that most borrowers repay their loans on time. To induce social pressure in the shame treatment (hereafter Shame), we send a message threatening to contact the endorsers if the borrower fails to repay on time. To create financial incentives without changing the budget conditions in the current period, we manipulate the interest rate for future loans. In the reward treatment (hereafter Reward), we send a message guaranteeing to reduce the interest rate for future borrowings by 5% if the borrower repays on time. In the punishment treatment (hereafter Punishment), we send a message threatening to increase the interest rate of future loans by 5% if the repayment is overdue. It is worth emphasizing that experimental interventions do not conflict with any clause in the borrowing contracts, and the platform implements all the incentives and penalties stated in the messages. Therefore, there is no deception in the experiment. In sum, we have the following treatments (the exact translation of the messages is provided in Appendix B.1):

Baseline ($n = 3768$: 1585 women and 2183 men): No message is sent.

Reminder ($n = 2823$: 1165 women and 1658 men): A simple reminder asking the participant to repay on time.

Norm ($n = 2807$: 1166 women and 1641 men): A message stating most borrowers make timely repayments and suggesting the borrower do the same.

Shame ($n = 2789$: 1161 women and 1628 men): A message stating the borrower's endorsers would be notified if the repayment is overdue.

Reward ($n = 2815$: 1172 women and 1643 men): A message promising to reduce the interest rate for future loans if the repayment is on time.

Punishment ($n = 2543$: 1022 women and 1521 men): A message stating the interest rate for future loans will increase if the repayment is overdue.

C. The procedures

The experiment was conducted between January 2017 and March 2017. During this period, borrowers with no overdue record were considered potential participants for our experiment, and they could only participate in one of the treatments at most. We excluded borrowers with overdue histories as they may have received other messages from the platform that could have contaminated their response to the incentives created by this experiment.[5] Any borrower who passed this screening process and had not repaid the loan one day before the due date was randomly placed into one of the treatment groups and received the corresponding message.

On 1 January 2017 (i.e., the first day of the experimental window), we identified 58,345 borrowers with no overdue record and due dates between 2nd January and 31st March. Of this group, 17,545 borrowers did not repay their loans by noon of the day before the corresponding deadline and therefore became participants in the experiment. The system randomly assigned each participant into one of the treatments; that is, the randomization was at the individual level.[6] At noon the day before the repayment deadline, each participant (except for the Baseline group) received a message that varied across treatments. To prevent other messages from confounding the experimental interventions, the platform did

---

[5]Prior to the experiment, the platform may call and/or text borrowers to remind them of impending deadlines if they had loans close to the corresponding due dates. The content of these communications was very much in the spirit of a reminder. However, passing the deadline point, the platform used a combination of different incentives, including but not limited to the incentives we used in the experiment, in the form of text messages, phone calls, and in-person visits, to demand repayment of the overdue loans.

[6]The original plan was to over-sample the baseline by 25% more observations than other treatments to enhance the statistical power for the comparisons between the baseline and treatments. However, due to a technical failure, the messages intended to be sent to the first batch of participants in the punishment treatment were not sent. Because this technical failure is independent of the characteristics of the participants, we pooled these participants into the baseline. The characteristics of the participants across treatments were balanced. Therefore, the failure is unlikely to invalidate our conclusions. Balance checks are reported in Appendix B.2.

not message the participants again unless the repayment was still not repaid 29 days later at which point, the loan went into default.[7]

Even though the platform applies the same standardized procedures to all borrowers in the borrowing process, it is unclear how it negotiates the interest rate with every first-time borrower. We find female borrowers pay higher interest rates than their male counterparts (see Table 1), and this difference persists after controlling for all other observable.[8] We also do not observe how the platform adjusts the interest rate after a borrower is overdue. These unobserved factors may confound the results by creating different incentive sizes (i.e., the interest rate) and setting different priors for men and women. However, these concerns are mitigated by the regression analyses, robustness checks, and experimental design. For the former two remedies, we control the interest rate variable in the regressions and perform the propensity score matching and the entropy balancing methods using all observed pre-origination demographic and credit characteristics; for the last remedy, we carefully excluded borrowers with overdue records because they have experienced the consequence and have a clear prior of what will happen past the deadline. It is also worth noting that the randomization is done at the individual level, making priors orthogonal to the assignment

[7]Our experimental design dictated no further communication with the experimental participants after the treatment intervention and before they had repaid or defaulted. However, the platform still made phone calls informing each of them about the loan in question and the consequences of defaulting after the payment was overdue. The content of the phone calls was not specific to the experiment and was independent of the treatments in the experiment. This caveat does not affect the borrowers' overdue choices as the platform made phone calls after loans were overdue, but it does distort the treatment effect measured by days overdue and default rate.

[8]There are several potential explanations for this gap. For example, female borrowers may be less inclined to negotiate the interest rate with the platform (Leibbrandt and List, 2015). The platform may also exercise double standard (Egan et al., 2022) and use algorithms that discriminate against women, as documented in the literature (Dobbie et al., 2021; Bartlett et al., 2022; Fuster et al., 2022).

of treatments.[9] There is also no known channel where borrowers share treatment-related information with each other. Therefore, these unobserved factors may not invalidate our findings.

D. Hypotheses

People suffer from limited attention (see, e.g., DellaVigna, 2009; Hirshleifer and Teoh, 2003) and are susceptible to nudges that evoke morality and pro-sociality (Allcott, 2011; Bursztyn et al., 2019b). Specifically, social norms (Hallsworth et al., 2017) and shame (Brocas et al., 2020; Kahan and Posner, 1999) have proven to be useful in encouraging compliance in other contexts. Consequently, a reminder message may reduce late repayment caused by forgetfulness, while messages that induce social pressure may reduce the severity of the moral hazard as the literature suggests. Thus, we expect all of our interventions to reduce the overdue rate of loans for participants.

Hypothesis 1: All of the treatments reduce the overdue rate as compared with the baseline.

So far, we have not taken gender differences into account. Hypotheses regarding the gender differences rely on the participants' pre-existing gender-specific characteristics (see Table 1), as well as documented gender differences in behaviors (i.e., gender differences in psychological responses).

[**Place Table 1 about here**]

For the Shame treatment, the effectiveness of the message may depend on the endorsers the borrower has chosen. Among experimental participants, female borrowers nominate 3.1 family members (1.9 friends) on average, whereas male borrowers only nominate 2.2 family members (2.8 friends). In Chinese culture, kinship is considered stronger than friendship. Thus, shame from a family member is likely to carry more weight than shame from a friend.

[9] Borrowers without any overdue record can still have priors from interactions with similar FinTech lenders before the experiment. The randomization of individuals also mitigates this concern.

Since the message sent to the Shame treatment group explicitly mentioned the possibility of contacting endorsers, we anticipate the impact of the choice of endorsers to be magnified. If so, women borrowers should be more responsive to the Shame treatment given they nominate more family members as endorsers. Furthermore, the literature finds that women tend to be more prone to shame (Ferguson et al., 2000; Lewis et al., 1992) and respond to shame more than men in other contexts such as stealing (Brocas et al., 2020). As a result, we predict that the Shame treatment is more effective for female borrowers as well.

For the Norm treatment, research in psychology and neuroscience shows that women are more sensitive than men to social cues determining what behavior is appropriate in certain contexts (see, e.g., Chen et al., 2019; Gilligan, 1993). Therefore, we hypothesize that female borrowers respond more than males to the Norm treatment as well.[10]

Hypothesis 2: Women respond more than men to the social incentives.

For the financial incentives, the magnitude of the reward/punishment depends on the product of the interest rate and the expected amount of borrowings in the future by design. Accordingly, the deterrence impact may depend on the interest rate and expectation of future borrowing behavior. If we use the current loan amount to proxy borrowers' credit needs in the future, the prediction is that men are more responsive to financial incentives as the product of the two terms is greater for men (men borrow 6.6% more than women at an interest rate 2.9% lower than women). We then turn to the psychological side and find several studies reveal that men are more sensitive to monetary losses and benefits than women (e.g., Kulich et al., 2011; Pokorny, 2008), although Bandiera et al. (2016) argue that this gender difference is small and study-specific.

In addition to gender differences in sensitivity to financial incentives, loss aversion (the phenomenon in which losses weigh more heavily than gains in decision processes) is another behavioral trait that may affect male and female borrowers differently. Since the founda-

[10]We address the impact of endorser choices and other psychological elements on the overdue rate in Section III.D.

tional work of Kahneman and Tversky (1979), researchers have found ample evidence that loss aversion is prevalent (Abdellaoui et al., 2007; Imas et al., 2017). Therefore, we expect Punishment to be more effective than Reward in influencing loan repayments for both genders. Besides, some studies find that men are more susceptible to loss aversion than women (Grolleau et al., 2016; Schmidt and Traub, 2002). As the preceding discussion suggests, both monetary and psychological arguments suggest that men are likely to respond more to the financial incentives.

Hypothesis 3: Both genders respond more to Punishment than Reward, and men respond more to financial incentives than women.

III. Results

We treat the overdue rate as the measure of non-compliance to study how the treatment interventions affect the borrowers' behavior. We use the overdue rate as the main outcome variable for several reasons. First, all of our treatments create incentives based on the repayment due date. Using the overdue rate is a fair comparison because the financial incentives perish after the deadline, while the social incentives may last longer. Second, having a low overdue rate is crucial to the development of the platform. Late repayments leave the affected lenders less willing to participate in the future, and the platform expends considerable resources reassuring the lenders who expect to reclaim their money on time and persuading the overdue borrowers not to default. Thus, it is in the platform's interest to reduce the overdue rate. Third, as mentioned earlier, due to the phone calls the platform made to the experimental participants who failed to repay on time, measuring the effect of incentives after the deadline may be contaminated. We report the results based on other measures of non-compliance, such as the duration of borrowers' overdue status and the default rate in Appendix C. The main findings are robust to different measurements.

Out of the 58,345 borrowers we observe during the experimental phase, 19,513 (66%)

male borrowers and 21,287 (75%) female borrowers repay at least one day early, with women more likely than men to repay early ($p = 0.0000$).[11] As we send messages only one day before the deadline, those who repay earlier do not receive the experimental interventions. To scrutinize the gender differences caused by the enforcement mechanisms, we focus on the 17,545 who entered the experiment in the following analyses.

A. Overdue rate in each treatment

Figure 2 plots the unconditional overdue rate across treatment groups. The messages significantly reduce the overdue rate from 37.1% in the Baseline to 34.2% in the Reminder treatment ($p = 0.0167$), 24.4% in the Norm treatment ($p < 0.0001$), 24.1% in the Shame treatment ($p < 0.0001$), 21.1% in the Reward treatment ($p < 0.0001$) and 17.0% in the Punishment treatment ($p < 0.0001$). The data confirm Hypothesis 1 that participants respond to both financial and non-financial incentives. The results also suggest that borrowers are more responsive to financial incentives (i.e., Reward or Punishment) than to social incentives (i.e., Norm or Shame), $p < 0.0001$. In summary, Punishment is the most effective intervention in the experiment where the impact is not disaggregated by gender.

[Place Figure 2 about here]

Result 1: The results confirm Hypothesis 1 that all of the treatments successfully reduce the overdue rate.

B. Gender differences in the unconditional overdue rate

Figure 3 decomposes the overdue rate by gender to see if the effects of the different treatments are affected by gender. As conjectured, responsiveness to the interventions is not gender-neutral. For women, the ranking of the mechanisms from the most effective to the least

[11] Without further specification, all p-values reported in the text are obtained from two-sided nonparametric Mann-Whitney tests.

effective is Shame (14.0% overdue), Norm (19.2% overdue), Punishment (22.2% overdue), Reward (27.2% overdue), and Reminder (37.2% overdue). In contrast, the ranking for men is Punishment (13.5% overdue), Reward (16.7% overdue), Norm (28.1% overdue), Shame (31.3% overdue), and Reminder (32.1% overdue). This suggests, if we know the genders of the participants, the most effective mechanism for female borrowers is Shame, while the most effective mechanism for male borrowers is Punishment. If we apply the most effective mechanism according to the gender of each participant, we can further reduce the overdue rate to 13.7%.

[**Place Figure 3 about here**]

C. Gender differences conditional on borrower characteristics

The above analyses focus on the overdue behavior unconditional on borrower characteristics. In Section II.D (Hypotheses), we conjecture that both gender differences in ex-ante characteristics and psychological factors may contribute to the observed behavior. In this section, we aim to control for borrower characteristics and quantify gender differences in the responsiveness to the enforcement mechanisms driven by psychological factors.[12] To achieve this goal, we regress the overdue rate on a set of dummy variables specifying treatments and genders, using all of the observed variables (as in Table 1) as controls. The regression is specified as following:

$$Overdue_i = \alpha + \sum_{j \in Treatments} \sum_{k \in Genders} \beta_{j,k} \mathbb{I}\{treatment_i = j\} \cdot \mathbb{I}\{gender_i = k\} + Controls_i + \epsilon_i,$$

where $Treatments$ is the set of all treatments in this study, $Genders$ is the set capturing the gender of participants, and $\mathbb{I}\{\cdot\}$ is an indicator that takes value 1 when the statement in the curly bracket is true and 0 otherwise. The control variables include age, the natural logarithm

[12]We discuss how ex-ante characteristics may contribute to the overall gender differences in Section III.D.

of monthly income, credit score (a categorical variable assigned by the platform taking integer values from 1 to 6, with 1 being the most credit-worthy), past borrowing incidences, loan amount, loan term, and loan interest rate of each borrower, as well as indicators for employment, higher education, marriage status, car ownership, home ownership, and holding loans from other financial intermediaries (i.e., other financial liabilities). Detailed definitions of these variables can be found in Table A.1 in Appendix A.1.

Table 2 shows the estimation results using the linear probability model, Logit model, and Probit model in Columns (1), (2), and (3), respectively. The sign and magnitude of the point estimates are consistent with the results from Figure 3. In the panel below the coefficients, we report the p-values testing the null hypotheses that the effectiveness of messages for each treatment group are the same for men and women compared with the baseline. For example, DID reminder tests $\mathcal{H}_0$: (Male reminder - Male baseline) = Female reminder.[13] The results suggest that the gender differences are robust to the control variables and estimation methods we consider.

[Place Table 2 about here]

C.1. Robustness checks

We then apply matching methods to check the robustness of the results above and further address the potential confound caused by pre-existing gender differences at the loan origination. Specifically, we apply the propensity-score matching method, which finds for each male borrower the female borrower with the closest observed characters, and then perform the same regression analyses as before on the matched sample. As King and Nielsen (2019) point out that propensity-score matching is restrictive in sample selection, we also report results using the entropy balancing method, which involves a re-weighting approach that provides

[13] Female baseline is omitted to avoid perfect multicollinearity in the regression, so it does not appear in the tests.

more flexibility than propensity-score matching (Hainmueller, 2012). Table 3 plots the estimation results. Columns (1)–(3) plot estimation results for the propensity-score matched sample, and Columns (4)–(6) show the results for the entropy-balanced sample. Although each methodology selects a different sample, all results found previously hold qualitatively, suggesting gender differences in responsiveness to enforcement mechanisms are robust to matching methods.

[Place Table 3 about here]

C.2. Heterogeneous analyses

The data finds strong evidence that, on average, women are more responsive to social pressure, while men are more sensitive to financial incentives. A natural question is whether borrowers' credit risk interacts with gender differences. The implication of this study will be different if high-risk or low-risk borrowers drive the gender differences. To estimate an individual's credit risk, we first regress the overdue dummy on all of the demographic, credit, and loan information in Table 1 using only the 3,768 observations from the Baseline treatment. The Logit model provides the following relationship:

$$\begin{aligned}
&\mathbb{E}(Overdue_i) = \Phi(0.488 - 0.275 Male_i + 0.006 Age_i - 0.006 Income_i \\
&- 0.231 Employ_i - 0.362 HighEdu_i + 0.244 Married_i + 0.245 Credit_i \\
&- 0.094 Car_i - 0.173 House_i - 0.129 OtherLoan_i - 0.013 PastIncidence_i \\
&- 0.110 LoanAmount_i - 0.01 LoanTerm_i - 0.006 IR_i - 0.048 Family_i).
\end{aligned}$$

Second, we estimate the probability of each borrower being overdue using the equation from the previous step and categorize each participant into the high-risk group if the estimated probability is greater than the median value and into the low-risk group otherwise.[14]

[14]The results are robust to different estimation methods (i.e., OLS and Probit/Logit) in Step 1 and different cut-offs separating the high- and low-risk groups in Step 2.

Table 4 displays the responsiveness to the enforcement mechanisms for both high-risk and low-risk groups. In Panel A, we perform regression analyses studying the deterrence impact of treatments for each sub-sample. While the data suggests that the deterrence effect of our interventions is more pronounced for the high-risk group (i.e., the absolute value of coefficients capturing treatment effects is larger for the high-risk group), the gender differences still hold. Panel B provides a finer look into the overdue rate for high- and low-risk groups by gender. In the Baseline, the overdue probability is 42% and 44% for high-risk men and women, respectively, while this drops to 33% and 20% in the Norm treatment. The same pattern holds for low-risk men and women, where the overdue probability reduces from 30% and 34% in the Baseline to 25% and 18% in the Norm treatment, respectively. Therefore, the gender differences apply to high-risk and low-risk borrowers alike.

[**Place Table 4 about here**]

Result 2: Women respond more to the social incentives. For female borrowers, Shame is the most effective mechanism to deter the breach of the repayment commitment.

Result 3: Men respond more to the financial incentives. For male borrowers, Punishment is the most effective mechanism to deter the breach of the repayment commitment.

D. Determinants of message effectiveness

In this section, we investigate potential channels, the choice of endorsers and (expected) size of financial incentives, that may explain the effectiveness of enforcement messages and shed light on the unconditional gender differences in the responsiveness to messages.

D.1. Endorser choice and Shame

As discussed in Section II.D (Hypotheses), the strength of the shame message conjectured to interact with the number of family members each borrower chooses as endorsers in affecting the overdue behavior. We examine how the choice of endorsers relates to the effectiveness of the shame message in this subsection.

We use the difference-in-differences (DID) approach to answer the question above, where we compare the changes in the overdue rate between Baseline/Reminder and Shame treatments for groups with more and fewer family member endorsers.[15] Concretely, we regress the overdue indicator on all the cross-terms generated by the dummy variable signifying whether a borrower nominates more family members than friends as endorsers (MoreFamily) and the dummy variable for the Shame treatment (Shame) as following:[16]

$$\begin{aligned} Overdue_i = \alpha + \beta_1 \mathbb{I}\{treatment_i = Shame\} + \beta_2 MoreFamily_i \\ + \beta_3 \mathbb{I}\{treatment_i = Shame\} \cdot MoreFamily_i + Controls_i + \epsilon_i. \end{aligned}$$

We start the analyses by regressing the overdue rate on MoreFamily and Shame respectively, and report results in Columns (1) and (2) of Table 5. The results show that having more family member endorsers and the presence of the shame message is correlated with a lower overdue rate. We then estimate the difference-in-differences (DID) specification by OLS, Logit, and Probit methods in Columns (3)–(5) to study the potential interaction effects between the choice of endorsers and the Shame treatment. First, the sign attached to the DID estimator (MoreFamily*Shame) is negative and significant in all specifications, suggesting that the shame message makes borrowers with more family member endorsers less likely to violate their contracts. Second, the sign of the Shame treatment dummy variable (Shame) is also negative and significant in all specifications, meaning that the shame treatment cause borrowers to be more obedient, disregarding the choices of endorsers.

As we find that the shame message cause borrowers to reduce overdue behavior through

[15] We excluded data for other treatments for a clean comparison; the results remain robust when we include observations for other treatments and control for the treatment fixed effects.

[16] We transform the number of family endorsers into a dummy variable (whether the number of family member endorsers exceeds half of all endorsers) in the regressions to make the interpretation easy to understand, and the results still hold if we treat it as a continuous or categorical variable.

the choice of endorsers and the direct impact of the message, a follow-up question would be whether the message produces the same effect size for men and women. To answer this question, we estimate the same regression model for men and women separately, and then compare the coefficients across regression models. Column (6) plots the results for women and Column (7) plots the results for men. First, from Column (6), we find women significantly respond to the shame message (Shame) and the interaction effect (MoreFamily*Shame), but men's response to either variable is not significant. Second, we perform t-tests to study whether these gender differences are significant across these two models (i.e., testing the triple difference estimator being zero). We find significant differences in both variables across equations; the shame message per se makes women less likely to be overdue than men with otherwise identical characteristics (-0.020 vs. -0.175, p =0.026). In addition, women with more family member endorsers are more responsive to the shame message (-0.015 vs. -0.088, p =0.071). To sum up, the gender difference in the responsiveness to the shame message can be explained by the gender difference in the ex-ante endorser choices (as shown in Table 1), the sensitively to the message per se, and the interaction between them.

However, it is still unclear why men and women react differently to the message and interaction effects from the above regressions because we do not observe the motivations behind endorser choices. There are several possibilities. First, men may be more strategic than women in selecting endorsers (i.e., men may nominate more endorsers who would not create much psychological discomfort upon default). Second, men and women may view the relative strength of kinship and friendship differently. Third, there may be gender differences in friend networks, and women do not have enough friends to choose from. This possibility is, however, unlikely to be the case as the literature finds that most people have more than five friends and women have more extensive friend networks than men (Gillespie et al., 2015). We encourage future experimental researchers to vary the composition of endorsers exogenously (i.e., randomize the quota for family endorsers) to disentangle these possibilities.

[**Place Table 5 about here**]

D.2. Expectation and financial incentives

In this subsection, we study whether the (expected) monetary size of punishment/reward affects the overdue behavior in the financial incentive treatments. From the experimental design, the expected reward or punishment size is proportional to the interest rate and the expected credit needs from the FinTech platform in the future. However, the latter is unobservable. To proxy the size of expected credit needs, we consider past borrowing incidences, past borrowing amount, and current loan amount because these variables capture borrowers' financial needs and reliance on the FinTech platform.

If borrowers respond to the size of financial incentives, then for two hypothetical borrowers with identical characteristics but placed in the Baseline and Reward/Punishment, the gap in the overdue probability between them widens as their interest rate or credit needs from the platform increase. Precisely, we compare the overdue choice in the two financial incentive treatments with those in the Baseline and Reminder, controlling for the level of (expected) size of financial incentives.

We perform regression analyses analogous to those in Section III.D.1 to examine whether these monetary considerations contribute to gender differences in borrowers' responsiveness to financial incentives. We regress the overdue indicator on all the cross-terms generated by the proxy for financial incentive size (Size), and the dummy variable for the financial incentive treatments (FinInctv) as the regression model below.[17]

$$
\begin{aligned}
Overdue_i = \alpha + \beta_1 \mathbb{I}\{treatment_i = FinInctv\} + \beta_2 Size_i \\
+ \beta_3 \mathbb{I}\{treatment_i = FinInctv\} \cdot Size_i + Controls_i + \epsilon_i.
\end{aligned}
$$

It is worth emphasizing that the objective of analyses here is not to use the incentive size to explain the overdue behavior per se, rather, these regressions use the incentive size to explain the difference in overdue behavior between the financial incentive treatments and other

[17] These results are qualitatively identical if we decompose the financial incentive dummy variable to the Reward and Punishment dummies, therefore, we combine the treatment in the main text for conciseness.

treatments. Therefore, the interpretation below focuses on the DID estimator (FinInctv*Size in Table 6).

[Place Table 6 about here]

Table 6 displays the estimation results of the regressions. Columns (1)–(4) report the estimation results when the size of the expected financial incentives is approximated by the interest rate, past borrowing incidences, past borrowing amount, and current loan amount, respectively.[18] We first focus on Column (1) when the size of financial incentives is approximated by the interest rate. We find that the direct impact of financial incentive messages significantly reduces the overdue rate (FinInctv<0). We also find that borrowers with higher interest rates are more sensitive to financial incentive treatments (FinInctv*Size<0), confirming our conjecture that the size of expected financial incentive can affect the effectiveness of messages providing financial incentives. However, when we turn the attention to the estimations for the variables approximating borrowers' future financial needs from the platform (Columns (2)–(4)), we only find the direct impact of financial incentive (FinInctv<0) is significant. The interaction term is no longer significant.[19]

We then employ the sub-sample analyses to study whether there are any gender differences in the above estimates. We estimate the same regression model as above by gender and report the results in Table 7. We report the corresponding results for men in Columns (1)–(4) and for women in Columns (5)–(8). We find that the results for either gender are qualitatively identical to the regression for both gender combined. We then test whether there are any gender differences in the coefficients. We find that the gender difference in the responsiveness to the financial incentive messages is mainly driven by the direct effect of the

[18]The results hold qualitatively true when we consider the size of financial incentives relative to borrowers' income levels.

[19]It is unclear why borrowers only respond to the interest rate but not proxies for future financial needs from the platform. One possibility would be that we use discount/markup of interest rate to create financial incentives. Therefore, the interest rate is more salient than other variables to borrowers.

existence of the financial incentive message, where the coefficient of FinInctv of men is significantly smaller than women for all four pairwise comparisons, while there is no significant gender difference in the coefficients for the interaction term.

To conclude, the gender differences in the responsiveness to financial incentives are likely to be driven by the gender difference in responsiveness to the existence of financial incentives rather than the size of financial incentives. As a limitation of this experiment, the expectations for future borrowing are endogenous and unobservable, therefore, it is difficult to accurately describe the relationship between the expected size of reward/punishment and the overdue behavior. Future researchers may solve these endogeneity concerns by randomly varying the size of financial incentives explicitly (i.e., experimenters could tell the participants that the compliance/non-compliance will lead to X RMB reduction/increase in the fee for the next loan, where the value of X is randomized).

[**Place Table 7 about here**]

IV. Discussion

We study how borrowers respond to different messages encouraging timely repayment in a FinTech lending context and explore the effectiveness of different interventions for men and women. Although we find that all of the interventions studied in the experiment successfully deter violations of financial commitments as expected, there are significant gender differences. Women are more responsive to social incentives, while men are more responsive to financial incentives. We then explore channels that may explain these gender differences. The data reveals that the gender differences in the responsiveness to the existence of shame message and the choice of endorsers can explain part of the overall gender difference for the Shame treatment. We also discover that the men are more responsive to the existence of financial incentives, but not to the size of these incentives, which explains the gender difference in the overall treatment effects. These results imply that the FinTech platform can customize the messages for male and female borrowers to reduce the delinquency rate. The data also suggest that some gender-neutral procedures and protocols, such as sending all borrowers the same message, may create unintended gender gaps in the financial market.

As some studies have discovered, gender differences may be context- and culture-specific (Andersen et al., 2018; Houser and Schunk, 2009; Wang et al., 2018). The extent to which we can apply these results is of critical interest to managers and regulators in the credit market, as well as policymakers aiming to discourage wrongdoers in general. We discuss the external validity of this study using the conditions proposed by List (2020). First, our sample is a random sample of borrowers having high default risks from the FinTech lending platform. Second, there is no attrition due to the nature of our experiment. Third, in our natural field experiment, participants do not know they are being observed; therefore, all participants make their decisions in a natural environment. As a result, we expect our findings to be valid for larger samples and borrowers from similar FinTech lending companies.

However, readers should proceed with caution in applying the same interventions to deter other types of non-compliance (Al-Ubaydli et al., 2017; Al-Ubaydli and List, 2015). Like most natural field experiments, our subjects pool may not represent the general public. We also do not observe borrowers who repaid early or had overdue records before the experiment.[20] More research is needed to generalize the findings to other setups. Despite the potential caveats above, our experiment provides a methodological tool to improve the overall effectiveness of enforcement mechanisms. Regulators are suggested to consider experimenting with various approaches to find the optimal way to customize enforcement mechanisms.

[20]We do find that there are some small but significant differences in the observed characteristics between those who repaid early and those who repaid late. Interested readers are referred to Appendix A.2 for a discussion about potential selection biases caused by the experimental design. However, as discussed, the threat of potential selection biases is moderated by our rigorous experimental design and analyses.

References

Abbink, Klaus, Utteeyo Dasgupta, Lata Gangadharan, and Tarun Jain, 2014, Letting the briber go free: An experiment on mitigating harassment bribes, *Journal of Public Economics* 111, 17–28.

Abdellaoui, Mohammed, Han Bleichrodt, and Corina Paraschiv, 2007, Loss aversion under prospect theory: A parameter-free measurement, *Management Science* 53, 1659–1674.

Al-Ubaydli, Omar, and John A List, 2015, On the generalizability of experimental results in economics, *Handbook of Experimental Economic Methodology* .

Al-Ubaydli, Omar, John A List, and Dana L Suskind, 2017, What can we learn from experiments? Understanding the threats to the scalability of experimental results, *American Economic Review* 107, 282–86.

Allcott, Hunt, 2011, Social norms and energy conservation, *Journal of Public Economics* 95, 1082–1095.

Andersen, Steffen, Seda Ertac, Uri Gneezy, John A List, and Sandra Maximiano, 2018, On the cultural basis of gender differences in negotiation, *Experimental Economics* 21, 757–778.

Andreoni, James, and Lise Vesterlund, 2001, Which is the fair sex? gender differences in altruism, *The Quarterly Journal of Economics* 116, 293–312.

Bandiera, Oriana, Greg Fischer, Andrea Prat, and Erina Ytsma, 2016, Do women respond less to performance pay? building evidence from multiple experiments, Working paper.

Bao, Zhengyang, and Difang Huang, 2021, Shadow banking in a crisis: Evidence from fintech during covid-19, *Journal of Financial and Quantitative Analysis* 56, 2320–2355.

Bao, Zhengyang, and Difang Huang, 2022, Reform scientific elections to improve gender equality, *Nature Human Behaviour* 6, 478–479.

Barber, Brad M, Wei Jiang, Adair Morse, Manju Puri, Heather Tookes, and Ingrid M Werner, 2021, What explains differences in finance research productivity during the pandemic?, *The Journal of Finance* 76, 1655–1697.

Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, 2022, Consumer-lending discrimination in the fintech era, *Journal of Financial Economics* 143, 30–56.

Brocas, Isabelle, Juan D Carrillo, and Mallory Montgomery, 2020, Shaming as an incentive mechanism against stealing: Behavioral and physiological evidence, *Journal of Public Economics* 194, 104351.

Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman, 2014, Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions, *Econometrica* 82, 1273–1301.

Bursztyn, Leonardo, Georgy Egorov, and Robert Jensen, 2019a, Cool to be smart or smart to be cool? Understanding peer pressure in education, *The Review of Economic Studies* 86, 1487–1526.

Bursztyn, Leonardo, Bruno Ferman, Stefano Fiorin, Martin Kanz, and Gautam Rao, 2018, Status goods: Experimental evidence from platinum credit cards, *The Quarterly Journal of Economics* 133, 1561–1595.

Bursztyn, Leonardo, Stefano Fiorin, Daniel Gottlieb, and Martin Kanz, 2019b, Moral incentives in credit card debt repayment: Evidence from a field experiment, *Journal of Political Economy* 127, 1641–1683.

Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais, 2017, 'Acting wife': Marriage market incentives and labor market investments, *American Economic Review* 107, 3288–3319.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott, 2020, Misperceived social norms: Women working outside the home in Saudi Arabia, *American Economic Review* 110, 2997–3029.

Bursztyn, Leonardo, and Robert Jensen, 2015, How does peer pressure affect educational investments?, *The Quarterly Journal of Economics* 130, 1329–1367.

Bursztyn, Leonardo, and Robert Jensen, 2017, Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure, *Annual Review of Economics* 9, 131–153.

Chen, Wanting, Shuyue Zhang, Ofir Turel, Youqing Peng, Hong Chen, and Qinghua He, 2019, Sex-based differences in right dorsolateral prefrontal cortex roles in fairness norm compliance, *Behavioural Brain Research* 361, 104–112.

Croson, Rachel, and Uri Gneezy, 2009, Gender differences in preferences, *Journal of Economic Literature* 47, 448–74.

de Roure, Calebe, Loriana Pelizzon, and Anjan Thakor, 2021, P2P Lenders versus Banks: Cream Skimming or Bottom Fishing?, *The Review of Corporate Finance Studies* 11, 213–262.

DellaVigna, Stefano, 2009, Psychology and economics: Evidence from the field, *Journal of Economic Literature* 47, 315–72.

DellaVigna, Stefano, John A List, and Ulrike Malmendier, 2012, Testing for altruism and social pressure in charitable giving, *The Quarterly Journal of Economics* 127, 1–56.

Di Maggio, Marco, and Vincent Yao, 2021, Fintech borrowers: Lax screening or cream-skimming?, *The Review of Financial Studies* 34, 4565–4618.

Dobbie, Will, Andres Liberman, Daniel Paravisini, and Vikram Pathania, 2021, Measuring bias in consumer lending, *The Review of Economic Studies* 88, 2799–2832.

Egan, Mark, Gregor Matvos, and Amit Seru, 2022, When Harry fired Sally: The double standard in punishing misconduct, *Journal of Political Economy* 130, 1184–1248.

Ferguson, Tamara J, Heidi L Eyre, and Michael Ashbaker, 2000, Unwanted identities: A key variable in shame–anger links and gender differences in shame, *Sex Roles* 42, 133–157.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther, 2022, Predictably unequal? the effects of machine learning on credit markets, *The Journal of Finance* 77, 5–47.

Gillespie, Brian Joseph, Janet Lever, David Frederick, and Tracy Royce, 2015, Close adult friendships, gender, and the life cycle, *Journal of Social and Personal Relationships* 32, 709–736.

Gilligan, Carol, 1993, *In a Different Voice Psychological Theory and Women's Development* (Harvard University Press).

Gneezy, Uri, and Aldo Rustichini, 2004, Gender and competition at a young age, *American Economic Review* 94, 377–381.

Grolleau, Gilles, Martin G Kocher, and Angela Sutan, 2016, Cheating and loss aversion: Do people cheat more to avoid a loss?, *Management Science* 62, 3428–3438.

Hainmueller, Jens, 2012, Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies, *Political Analysis* 20, 2546.

Hallsworth, Michael, John A List, Robert D Metcalfe, and Ivo Vlaev, 2017, The behavioralist as tax collector: Using natural field experiments to enhance tax compliance, *Journal of Public Economics* 148, 14–31.

Harrison, Glenn W, and John A List, 2004, Field experiments, *Journal of Economic literature* 42, 1009–1055.

Hirshleifer, David, and Siew Hong Teoh, 2003, Limited attention, information disclosure, and financial reporting, *Journal of Accounting and Economics* 36, 337–386.

Houser, Daniel, and Daniel Schunk, 2009, Social environments with competitive pressure: Gender effects in the decisions of german schoolchildren, *Journal of Economic Psychology* 30, 634–641.

Imas, Alex, Sally Sadoff, and Anya Samek, 2017, Do people anticipate loss aversion?, *Management Science* 63, 1271–1284.

Kahan, Dan M, and Eric A Posner, 1999, Shaming white-collar criminals: A proposal for reform of the federal sentencing guidelines, *The Journal of Law and Economics* 42, 365–392.

Kahneman, Daniel, and Amos Tversky, 1979, Prospect theory: an analysis of decision under risk, *Econometrica* 47, 263–291.

Karlan, Dean, Melanie Morten, and Jonathan Zinman, 2012, A personal touch: Text messaging for loan repayment, Technical report, National Bureau of Economic Research.

King, Gary, and Richard Nielsen, 2019, Why propensity scores should not be used for matching, *Political Analysis* 27, 435–454.

Kulich, Clara, Grzegorz Trojanowski, Michelle K Ryan, S Alexander Haslam, and Luc DR Renneboog, 2011, Who gets the carrot and who gets the stick? evidence of gender disparities in executive remuneration, *Strategic Management Journal* 32, 301–321.

Leibbrandt, Andreas, and John A List, 2015, Do women avoid salary negotiations? evidence from a large-scale natural field experiment, *Management Science* 61, 2016–2024.

Lewis, Michael, Steven M Alessandri, and Margaret W Sullivan, 1992, Differences in shame and pride as a function of children's gender and task difficulty, *Child development* 63, 630–638.

List, John A, 2020, Non est disputandum de generalizability? a glimpse into the external validity trial, Working paper.

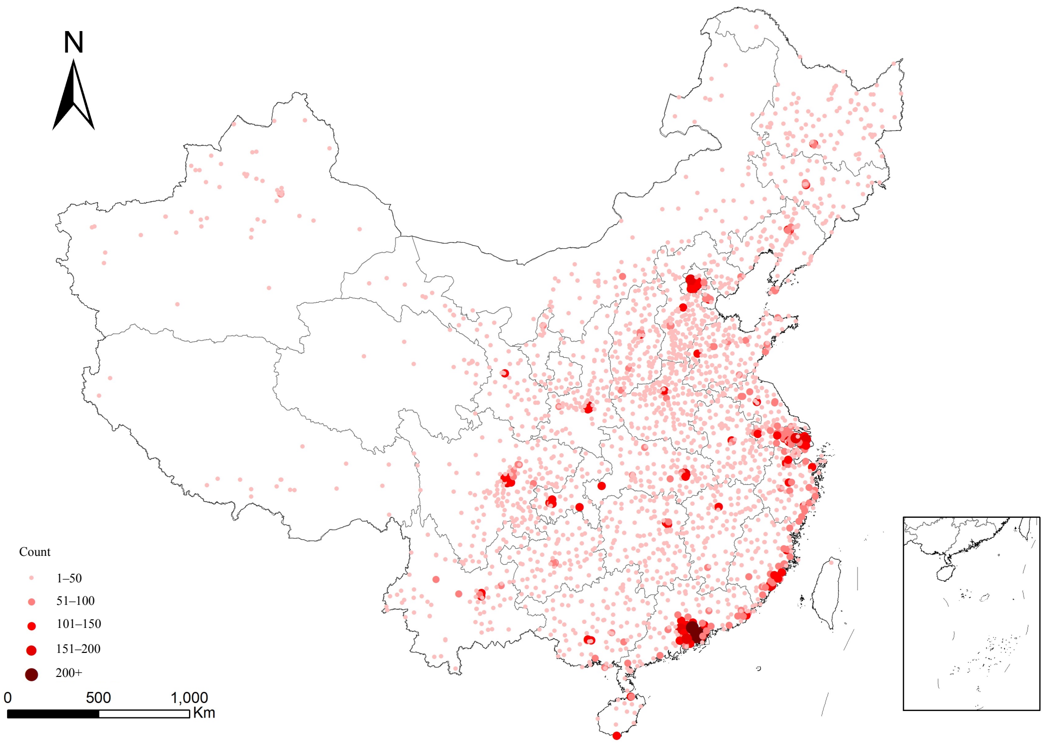
Niederle, Muriel, and Lise Vesterlund, 2007, Do women shy away from competition? do men compete too much?, *The Quarterly Journal of Economics* 122, 1067–1101.

Pokorny, Kathrin, 2008, Pay–but do not pay too much: An experimental study on the impact of incentives, *Journal of Economic Behavior & Organization* 66, 251–264.

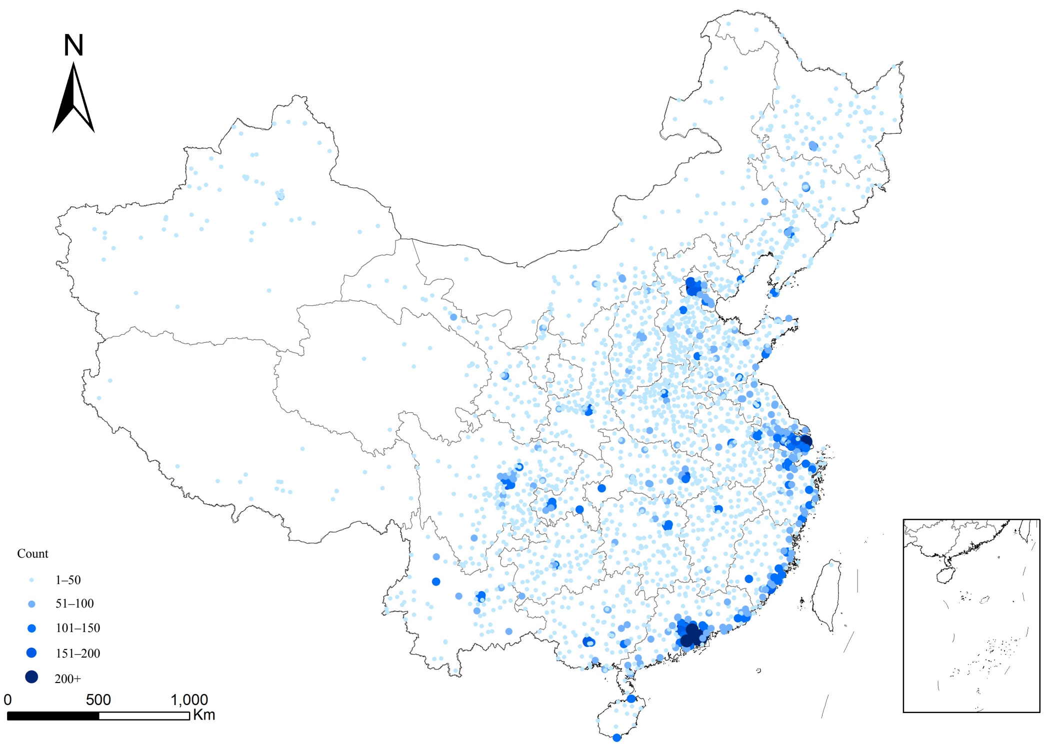
Schmidt, Ulrich, and Stefan Traub, 2002, An experimental test of loss aversion, *Journal of Risk and Uncertainty* 25, 233–249.

Wang, Jianxin, Daniel Houser, and Hui Xu, 2018, Culture, gender and asset prices: Experimental evidence from the us and china, *Journal of Economic Behavior & Organization* 155, 253–287.

Xiao, Erte, and Daniel Houser, 2011, Punish in public, *Journal of Public Economics* 95, 1006–1017.



(a) Location of female borrowers



(b) Location of male borrowers

Figure 1: Geographic location of experiment borrowers. This figure presents the geographical distribution of experiment participants at the city level for female borrowers in Panel (a) and male borrowers in Panel (b).

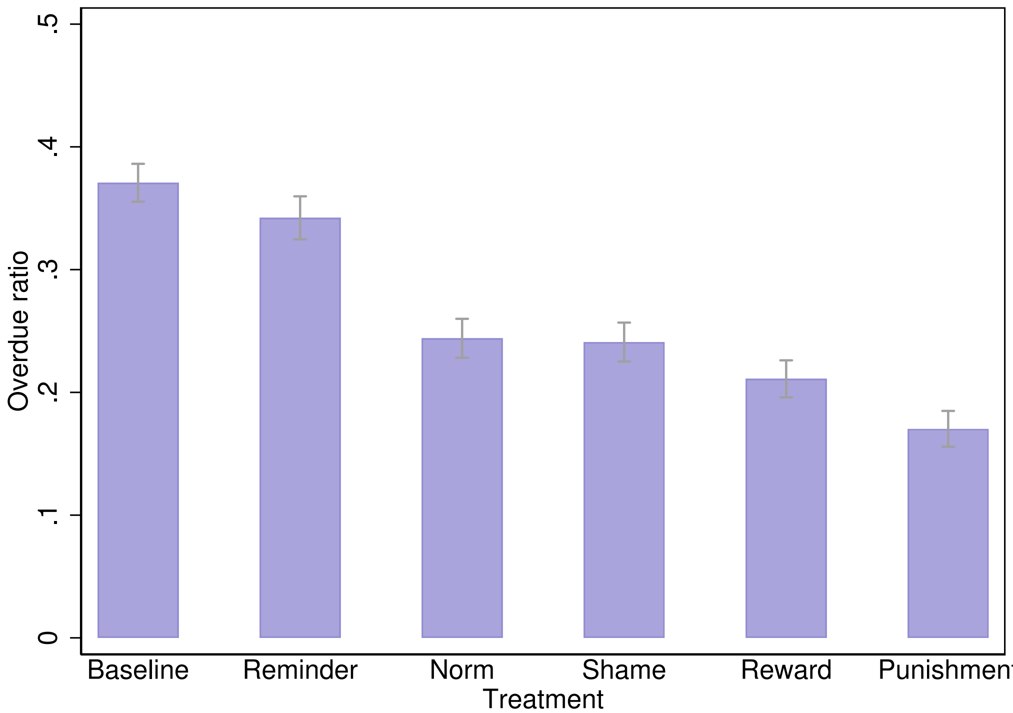


Figure 2: Overdue ratio. This figure plots the overdue ratio across treatments. The error bars plot the 95% confidence interval.

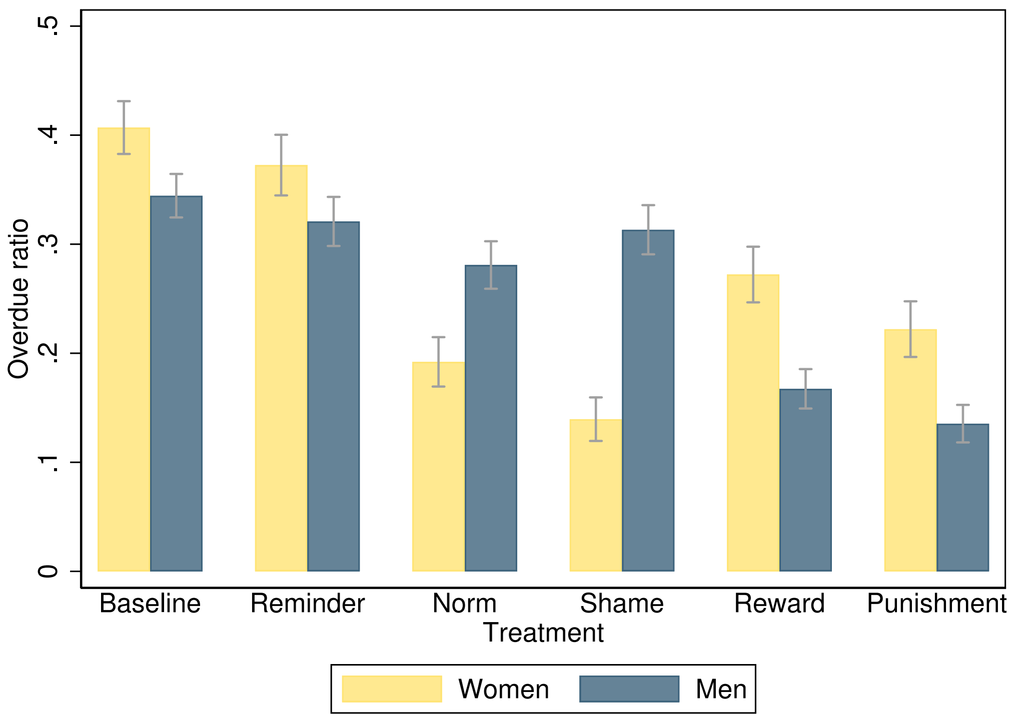


Figure 3: Overdue ratio by gender. This figure plots the overdue ratio across treatments by gender. The error bars plot the 95% confidence interval.

Table 1: Gender differences in characteristics. This table presents the pre-existing gender differences in demographic, credit, and loan characteristics for all 58,345 borrowers we observed. We report the mean value for each gender and the corresponding gender difference. *, **, and * indicate statistical significance (of difference in means tests for men and women samples) at the 10%, 5%, and 1% levels, respectively.**

Variables	Unit	Men	Women	Difference	*p*-value
Panel A: Demographic Variables					
Age	years	28.942	28.545	0.396***	0.000
Monthly Income	RMB	4585.29	3550.96	1034.33***	0.000
Employment Indicator	0/1	0.585	0.585	0.000	0.982
High Education Indicator	0/1	0.679	0.680	-0.001	0.861
Married Indicator	0/1	0.490	0.493	-0.003	0.516
Panel B: Credit Variables					
Credit Score	1–6	2.811	2.868	-0.057***	0.000
Car Indicator	0/1	0.447	0.443	0.004	0.337
House Indicator	0/1	0.616	0.611	0.004	0.277
Other Loan Indicator	0/1	0.208	0.211	-0.003	0.331
Overdue Record Indicator	0/1	0.000	0.000	–	–
Past Borrowing Incidence	times	2.147	2.252	-0.105***	0.000
Historical Loan Amount	RMB	13690.01	13041.22	648.80	0.274
Panel C: Loan Information Variables					
Loan Amount	RMB	8502.89	7972.96	529.93***	0.000
Loan Term	months	9.669	9.242	0.426***	0.000
Interest Rate	%	16.385	16.880	-0.495***	0.000
Family Endorsers	0–5	2.215	3.148	-0.932***	0.000
Sample Size	–	29,787	28,558	–	–

Table 2: Regressions comparing changes in the overdue rate across treatments by gender. This table reports OLS, Logit, and Probit estimates of the overdue rate under different treatments by gender. t-statistics are reported in parentheses below each coefficient. *, **, and * indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The bottom panel displays the p-values testing the null hypotheses that both genders reduce their overdue rate at the same magnitude in response to the treatment as compared with the baseline.**

	Dependent Variable: Overdue Indicator		
	(1)	(2)	(3)
Male baseline	-0.063***	-0.276***	-0.169***
	(-3.89)	(-3.86)	(-3.86)
Male reminder	-0.088***	-0.393***	-0.239***
	(-5.16)	(-5.12)	(-5.11)
Male norm	-0.129***	-0.595***	-0.362***
	(-7.72)	(-7.61)	(-7.64)
Male shame	-0.099***	-0.445***	-0.274***
	(-5.81)	(-5.78)	(-5.83)
Male reward	-0.242***	-1.266***	-0.748***
	(-15.44)	(-14.58)	(-14.83)
Male punish	-0.270***	-1.499***	-0.875***
	(-17.55)	(-16.00)	(-16.46)
Female reminder	-0.026	-0.107	-0.067
	(-1.42)	(-1.33)	(-1.35)
Female norm	-0.204***	-1.025***	-0.610***
	(-12.15)	(-11.23)	(-11.44)
Female shame	-0.257***	-1.415***	-0.833***
	(-16.28)	(-14.28)	(-14.88)
Female reward	-0.128***	-0.586***	-0.355***
	(-7.21)	(-6.95)	(-6.97)
Female punish	-0.182***	-0.879***	-0.528***
	(-10.26)	(-9.58)	(-9.71)
Controls	Yes	Yes	Yes
Method	OLS	Logit	Probit
N	17,545	17,545	17,545
DID reminder	p =0.951	p =0.921	p =0.961
DID norm	p =0.000	p =0.000	p =0.000
DID shame	p =0.000	p =0.000	p =0.000
DID reward	p =0.024	p =0.001	p =0.001
DID punish	p =0.253	p =0.007	p =0.016

Table 3: Regressions comparing overdue rates across treatments based on matched samples. This table reports OLS, Logit, and Probit estimates of the overdue rate under different treatments by gender in two matched samples. Columns (1)–(3) report results using the matched sample based on the propensity score method, and Columns (4)–(6) show results using the matched sample based on the entropy balancing method. t-statistics are reported in parentheses below each coefficient. *, **, and * indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The bottom panel displays the p-values testing the null hypotheses that both genders reduce their overdue rate at the same magnitude in response to the treatment as compared with the baseline.**

	Dependent Variable: Overdue Indicator					
	Propensity Score Matched Sample			Entropy Balanced Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
Male baseline	-0.052*	-0.212*	-0.130*	-0.062***	-0.259***	-0.159***
	(-1.91)	(-1.83)	(-1.82)	(-2.98)	(-2.88)	(-2.87)
Male reminder	-0.077***	-0.331***	-0.201***	-0.087***	-0.377***	-0.230***
	(-2.79)	(-2.78)	(-2.73)	(-4.07)	(-4.01)	(-3.98)
Male norm	-0.118***	-0.535***	-0.326***	-0.129***	-0.578***	-0.353***
	(-4.32)	(-4.43)	(-4.39)	(-6.03)	(-6.05)	(-6.04)
Male shame	-0.088***	-0.382***	-0.236***	-0.098***	-0.428***	-0.265***
	(-3.17)	(-3.18)	(-3.18)	(-4.54)	(-4.51)	(-4.54)
Male reward	-0.231***	-1.206***	-0.711***	-0.241***	-1.250***	-0.739***
	(-8.60)	(-9.53)	(-9.33)	(-11.76)	(-12.16)	(-12.13)
Male punish	-0.259***	-1.438***	-0.837***	-0.270***	-1.482***	-0.865***
	(-9.73)	(-10.96)	(-10.74)	(-13.29)	(-13.66)	(-13.72)
Female reminder	-0.007	-0.022	-0.014	-0.038	-0.157	-0.098
	(-0.20)	(-0.14)	(-0.15)	(-1.39)	(-1.33)	(-1.35)
Female norm	-0.173***	-0.836***	-0.501***	-0.202***	-0.990***	-0.591***
	(-4.87)	(-4.55)	(-4.61)	(-8.28)	(-7.73)	(-7.85)
Female shame	-0.208***	-1.045***	-0.626***	-0.229***	-1.161***	-0.693***
	(-5.96)	(-5.46)	(-5.60)	(-9.00)	(-8.01)	(-8.32)
Female reward	-0.121***	-0.549***	-0.335***	-0.118***	-0.527***	-0.323***
	(-3.28)	(-3.15)	(-3.18)	(-4.48)	(-4.32)	(-4.37)
Female punish	-0.205***	-1.019***	-0.605***	-0.205***	-0.995***	-0.595***
	(-5.86)	(-5.38)	(-5.44)	(-8.40)	(-7.86)	(-7.97)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Method	OLS	Logit	Probit	OLS	Logit	Probit
N	13,493	13,493	13,493	17,545	17,545	17,545
DID reminder	p =0.657	p =0.575	p =0.598	p =0.680	p =0.779	p =0.743
DID norm	p =0.006	p =0.001	p =0.009	p =0.000	p =0.000	p =0.000
DID shame	p =0.000	p =0.000	p =0.000	p =0.000	p =0.000	p =0.000
DID reward	p =0.139	p =0.021	p =0.033	p =0.041	p =0.001	p =0.003
DID punish	p =0.094	p =0.032	p =0.040	p =0.093	p =0.014	p = 0.021

Table 4: Borrower quality and overdue frequency. This table shows how high- and low-credit risk borrowers respond to enforcement mechanisms by gender. In Panel (a), we report OLS, Logit, and Probit estimates of the overdue rate under different treatments by gender for low-risk borrowers in Columns (1)–(3) and high-risk borrowers in Columns (4)–(6). *t*-statistics are reported in parentheses below each coefficient. *, **, and * indicate statistical significance at the 10%, 5%, and 1% levels, respectively. In Panel (b), we show the overdue likelihood of borrowers in the corresponding category for each treatment by gender.**

(a) Estimates of borrowers' overdue rate under different treatments

	Dependent Variable: Overdue Indicator					
	Low-Risk Borrowers			High-Risk Borrowers		
	(1)	(2)	(3)	(4)	(5)	(6)
Male baseline	-0.067***	-0.310***	-0.188***	-0.062**	-0.314***	-0.185***
	(-2.88)	(-3.10)	(-3.04)	(-2.49)	(-2.70)	(-2.63)
Male reminder	-0.141***	-0.632***	-0.385***	-0.055**	-0.282**	-0.166**
	(-5.76)	(-5.80)	(-5.78)	(-2.12)	(-2.34)	(-2.27)
Male norm	-0.160***	-0.719***	-0.439***	-0.111***	-0.567***	-0.337***
	(-6.54)	(-6.48)	(-6.48)	(-4.37)	(-4.57)	(-4.52)
Male shame	-0.090***	-0.409***	-0.250***	-0.107***	-0.546***	-0.326***
	(-3.68)	(-3.84)	(-3.80)	(-4.18)	(-4.36)	(-4.33)
Male reward	-0.304***	-1.520***	-0.907***	-0.199***	-1.127***	-0.658***
	(-13.82)	(-12.21)	(-12.59)	(-8.21)	(-8.47)	(-8.48)
Male punish	-0.320***	-1.630***	-0.969***	-0.238***	-1.455***	-0.832***
	(-14.38)	(-12.12)	(-12.62)	(-9.95)	(-10.27)	(-10.35)
Female reminder	-0.017	-0.068	-0.042	-0.043	-0.196	-0.118
	(-0.72)	(-0.69)	(-0.68)	(-1.43)	(-1.36)	(-1.36)
Female norm	-0.232***	-1.133***	-0.676***	-0.156***	-0.850***	-0.504***
	(-10.83)	(-9.91)	(-10.10)	(-5.79)	(-5.53)	(-5.64)
Female shame	-0.276***	-1.454***	-0.864***	-0.217***	-1.348***	-0.775***
	(-13.83)	(-11.95)	(-12.44)	(-8.47)	(-7.60)	(-7.91)
Female reward	-0.146***	-0.656***	-0.399***	-0.089***	-0.441***	-0.261***
	(-6.60)	(-6.36)	(-6.39)	(-3.01)	(-2.97)	(-2.95)
Female punish	-0.208***	-0.992***	-0.600***	-0.123***	-0.631***	-0.374***
	(-9.61)	(-8.85)	(-9.03)	(-4.05)	(-3.89)	(-3.91)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Method	OLS	Logit	Probit	OLS	Logit	Probit
N	8,772	8,772	8,772	8,773	8,773	8,773

(b) Average borrowers' overdue rate under different treatments

	Male	Female	Male	Female	Male	Female
	Baseline		Reminder		Norm	
High-Risk Borrowers	0.422	0.441	0.349	0.421	0.331	0.203
Low-Risk Borrowers	0.300	0.337	0.302	0.286	0.248	0.176
	Shame		Reward		Punish	
High-Risk Borrowers	0.398	0.154	0.181	0.286	0.162	0.224
Low-Risk Borrowers	0.253	0.116	0.158	0.246	0.119	0.217

Table 5: Choice of endorsers and overdue rate. This table reports estimates of the relationship between the overdue rate and the number of family endorsers in the Shame treatment. *t*-statistics are reported in parentheses below each coefficient. *, **, and * indicate statistical significance at the 10%, 5%, and 1% levels, respectively.**

	Dependent Variable: Overdue Indicator						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MoreFamily	-0.038***		-0.005	-0.020	-0.009	-0.027*	-0.024
	(-3.67)		(-0.38)	(-0.36)	(-0.26)	(-1.78)	(-1.09)
Shame		-0.116***	-0.051***	-0.236***	-0.146***	-0.020	-0.175***
		(-11.77)	(-3.29)	(-3.24)	(-3.30)	(-1.13)	(-5.36)
MoreFamily*Shame			-0.119***	-0.670***	-0.392***	-0.015	-0.088**
			(-5.94)	(-6.39)	(-6.32)	(-0.53)	(-2.46)
Constant	0.482***	0.468***	0.485***	0.008	-0.024	0.504***	0.398**
	(-4.64)	(-4.54)	(-4.70)	(-0.02)	(-0.08)	(-3.62)	(-2.53)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Method	OLS	OLS	OLS	Logit	Probit	OLS	OLS
Sample	Full	Full	Full	Full	Full	Male	Female
N	9,380	9,380	9,380	9,380	9,380	5,469	3,911

Table 6: Size of financial incentives and overdue rate. This table explores the relationship between expected financial incentive size with overdue behavior. t-statistics are reported in parentheses below each coefficient. *, **, and * indicate statistical significance at the 10%, 5%, and 1% levels, respectively.**

	Dependent Variable: Overdue Indicator			
Size is:	Loan Interest Rate	Past Borrowing Incidence	Historical Loan Amount	Current Loan Amount
	(1)	(2)	(3)	(4)
Size	0.039**	-0.036**	-0.007	-0.004
	(-2.23)	(-2.15)	(-0.58)	(-0.29)
FinInctv	-0.150***	-0.178***	-0.175***	-0.174***
	(-16.67)	(-13.39)	(-15.26)	(-15.61)
Size*FinInctv	-0.079***	0.020	0.020	0.019
	(-3.86)	(1.19)	(1.24)	(1.17)
Constant	0.572***	0.472***	0.582***	0.582***
	(6.52)	(4.42)	(6.62)	(6.62)
Controls	Yes	Yes	Yes	Yes
Method	OLS	OLS	OLS	OLS
N	11,949	11,949	11,949	11,949

Table 7: Gender differences in responding to financial incentives. This table reports estimates of the relationship between the overdue rate and the expected size of financial incentive by gender. t-statistics are reported in parentheses below each coefficient. *, **, and * indicate statistical significance at the 10%, 5%, and 1% levels, respectively.**

	Dependent Variable: Overdue Indicator							
	Male				Female			
Size is:	Loan Interest Rate	Past Borrowing Incidence	Historical Loan Amount	Current Loan Amount	Loan Interest Rate	Past Borrowing Incidence	Historical Loan Amount	Current Loan Amount
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Size	0.033	-0.020	-0.008	-0.007	0.049*	-0.056*	-0.008	-0.002
	(1.52)	(-0.95)	(-0.51)	(-0.38)	(1.68)	(-1.92)	(-0.39)	(-0.08)
FinInctv	-0.169***	-0.190***	-0.196***	-0.195***	-0.124***	-0.154***	-0.146***	-0.144***
	(-14.80)	(-12.17)	(-13.52)	(-13.85)	(-8.54)	(-6.32)	(-7.82)	(-7.93)
Size*FinInctv	-0.068***	0.013	0.027	0.027	-0.091***	0.019	0.011	0.007
	(-2.68)	(0.63)	(1.35)	(1.33)	(-2.64)	(0.65)	(0.40)	(0.25)
Constant	0.421***	0.378***	0.432***	0.434***	0.638***	0.435**	0.644***	0.641***
	(3.76)	(2.85)	(3.86)	(3.88)	(4.32)	(2.32)	(4.35)	(4.33)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Method	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
N	7,005	7,005	7,005	7,005	4,944	4,944	4,944	4,944

Appendix

Appendix A displays further information on the FinTech platform; Appendix B provides further details on the experiment, and Appendix C provides further analysis on days overdue and default likelihood.

A. Further platform information

A.1. Variables and definitions

We define all variables mentioned in the paper in Table A.1.

A.2. Borrower characteristics and potential selection bias

Two design choices may lead the participants who received the treatment interventions not to represent the overall borrower population using the platform. First, we conducted the experiment in the first quarter of the year. This may be problematic if borrower characteristics are seasonal. Second, the exclusion of borrowers with overdue histories may cause the sample to be not representative of the full population of borrowers. To address these selection issues and to understand the characteristics of a typical borrower, we obtained the entire sample of borrowers holding credit loans with the same amortization scheme terminating during the year 2016 (i.e., the year before the experiment). We refer to this sample as the Yearly sample in contrast to the 58,345 borrowers we collected for the experimental purpose (hereafter the Experimental sample) and compare them closely in this section.

First, we study the characteristics of the borrowers in our samples, especially the pre-existing gender differences. Table A.2 depicts the means of key borrower characteristics by gender. In the experimental sample, gender is roughly balanced (51% men, 49% women); men are 28.9 years old, 0.4 years older than women ($p = 0.0000$); and men earn RMB 4,585 per month, 1,034 more than women ($p = 0.0000$). Both men and women have a 59% chance

Table A.1: Variable definition. This table defines all variables used in this paper, including the demographic variables in Panel A, the credit variables in Panel B, and the loan information variables in Panel C.

Variable	Definition
	Panel A: Demographic Variables
Gender (Male)	Indicator that equals one if the borrower is male.
Age	Age of the borrower.
Monthly Income	The borrower's monthly income.
Employment Indicator	Indicator that equals one if the borrower is currently employed and zero otherwise.
Higher Education Indicator	Indicator that equals one if the borrower has received a bachelor's degree or above and zero otherwise.
Married Indicator	Indicator that equals one if the borrower is currently married and zero otherwise.
	Panel B: Credit Variables
Credit Score	The platform assigns each borrower a score on a scale of 1 to 6 with 1 being the most credit-worthy.
Car Indicator	Indicator that equals one if the borrower's household owns a car.
House Indicator	Indicator that equals one if the borrower's household owns a piece of real estate.
Other Loan Indicator	Indicator that equals one if the borrower has other loans from banks or other financial intermediaries.
Overdue Record Indicator	Indicator that equals one if the borrower has an overdue record at the platform.
Past Borrowing Incidence	Number of borrowing incidences for each borrower.
Historical Loan Amount	Historical borrowing amount in RMB for each borrower.
	Panel C: Loan Information Variables
Loan Amount	Loan amount in RMB.
Loan Term	The maturity of the loan.
Interest Rate	Annualized interest rate.
Family Endorsers	Number of family members the borrower chose as endorsers.
Overdue Indicator	Indicator that equals one if the loan is not paid back before the deadline.
Default Indicator	Indicator that equals one if the loan is not paid back before the expiry of the 29-day window after the deadline.

of being employed, a 68% chance of holding a university degree, and a 49% chance of being married, and there is no significant gender difference for these variables. For the credit variables, we find that men's credit score (a categorical variable assigned by the platform taking integer values from 1 to 6 with 1 being the most credit-worthy) is 0.06 lower (i.e., better) than women's on average ($p = 0.0000$), and men have borrowed 0.1 times less than women in the past ($p = 0.0000$). There is no significant gender difference in other credit variables such as having a car, property or loans from other financial intermediaries. For the loan variables, men borrow RMB 530 more ($p = 0.0000$) with 0.4 months longer in duration ($p = 0.0000$) than women. Moreover, the interest rate for men is 0.5 percentage points lower than that of women ($p = 0.0000$), and men nominate 0.9 fewer family members as endorsers than women ($p = 0.0000$).

Table A.2: Borrower characteristics. This table presents demographic, credit, and loan characteristics for the borrowers in the Experimental and Yearly samples. We report the mean value for each gender and the corresponding gender difference. *, **, and * indicate statistical significance (of difference in means tests for men and women borrowers within each sample) at the 10%, 5%, and 1% levels, respectively.**

	Experimental sample			Yearly sample			Yearly sample without overdue history			Yearly sample of first quarter		
	Men	Women	Diff	Men	Women	Diff	Men	Women	Diff	Men	Women	Diff
Panel A: Demographic Variables												
Gender (male=1)	0.511		–	0.504		–	0.508		–	0.505		–
Age	28.942	28.545	0.396***	28.932	28.583	0.349***	28.802	28.461	0.341***	28.923	28.538	0.385***
Monthly Income	4585.29	3550.96	1034.33***	4288.16	3249.09	1039.07***	4331.34	3283.01	1048.33***	4257.97	3238.21	1019.76***
Employment Indicator	0.585	0.585	0.000	0.575	0.575	-0.001	0.583	0.581	0.002	0.578	0.576	0.002
Higher Education Indicator	0.679	0.680	-0.001	0.668	0.683	-0.015*	0.676	0.690	-0.014*	0.669	0.682	-0.013*
Married Indicator	0.490	0.493	-0.003	0.490	0.490	0.000	0.489	0.489	0.000	0.490	0.491	-0.001
Panel B: Credit Variables												
Credit Score	2.811	2.868	-0.057***	2.830	2.888	-0.058***	2.797	2.894	-0.097***	2.829	2.883	-0.053***
Car Indicator	0.447	0.443	0.004	0.441	0.441	0.000	0.450	0.447	0.003	0.440	0.436	0.003
House Indicator	0.616	0.611	0.004	0.614	0.600	0.014*	0.623	0.601	0.022**	0.614	0.600	0.014*
Other Loan Indicator	0.208	0.211	-0.003	0.207	0.212	-0.005	0.206	0.210	-0.005	0.205	0.211	-0.006
Overdue Record Indicator	0.000	0.000	0.000	0.136	0.149	-0.013	0.000	0.000	0.000	0.136	0.149	-0.012
Past Borrowing Incidence	2.147	2.252	-0.105***	2.453	2.473	-0.020***	2.177	2.186	-0.009***	2.457	2.493	-0.036***
Historical Loan Amount	13690.01	13041.22	648.80	15868.04	15247.11	620.92	15491.72	15004.72	487.00	16544.67	15455.93	1088.74
Panel C: Loan Information Variables												
Loan Amount	8502.89	7972.96	529.93***	5746.91	5626.13	120.78***	5821.06	5755.53	65.54***	5810.53	5610.02	200.51***
Loan Term	9.669	9.242	0.426***	9.589	9.468	0.122***	9.535	9.412	0.123***	9.581	9.467	0.113***
Loan Interest Rate (%)	16.385	16.880	-0.495***	16.836	16.961	-0.125***	16.673	16.883	-0.210***	16.839	16.941	-0.101***
Family Endorsers	2.215	3.148	-0.932***	2.222	3.141	-0.919***	2.225	3.142	-0.917***	2.212	3.146	-0.934***
Overdue Rate (%)	9.185	7.042	2.143***	13.259	10.966	2.293***	12.045	9.734	2.311***	12.644	10.645	2.000***
Default Rate (%)	1.269	1.422	-0.153***	3.704	3.165	0.539***	3.142	2.592	0.551***	3.633	3.018	0.615***
Sample Size	29,787	28,558	–	112,348	110,619	–	97,096	94,149	–	29,673	29,094	–

Some statistically significant discrepancies emerge when we compare the Experimental sample with the Yearly sample. For example, we find that the gender ratio is 0.7 percentage points larger in the Experimental sample ($p = 0.0042$). Borrowers in the Experimental sample also have a higher monthly income, a larger probability of being employed, and a better chance of having higher education. They also borrow larger loan amounts; however, they have fewer borrowing incidences in the past and have no overdue records, as the experimental design mechanically removes borrowers with an overdue history ($p = 0.0000$ for all comparisons in this and the previous sentence). Importantly, all statistically significant gender differences in the Experimental sample are preserved in the Yearly sample.

To study the aforementioned selection concerns, we divide the Yearly sample into two sub-samples: those without overdue records and those with due dates in the first quarter, and we analyze these sub-samples in Table A.2. We first compare the sub-sample without overdue records with the Experimental sample and find that most of the differences between the Yearly and Experimental samples resolve after excluding the borrowers with an overdue record from the Yearly sample. Except for the borrowers' income and loan amount, the Experimental sample exhibits no statistical differences from the sub-sample. We then look into the sub-sample of Yearly with loan deadlines in the first quarter of 2016. This sub-sample is very similar to the full Yearly sample, except that the borrowers have lower wages ($p = 0.0383$). To conclude, the discrepancies between the Experimental and Yearly samples are driven mainly by the exclusion of the borrowers with overdue records, and the borrowers with loans ending in the first quarter are representative of the general borrower population.

These comparisons suggest that the potential selection bias is unlikely to undermine the inferences from our experiment. First, we find that the borrowing records terminating in the first quarter are representative of those in the Yearly sample. The fact that borrowers in the Yearly sample have a slightly higher wage than those in the first quarter sub-sample is likely due to inflation and economic development. Second, dropping borrowers with overdue records does indeed causes some variables to change, but these variations, while highly statistically significant, are small in magnitude (most are within 1% of the value from the Yearly sample). Additionally, borrowers with overdue histories account for only 14% of the Yearly population, and there is no evidence suggesting that having overdue records should change the relative responsiveness to enforcement mechanisms between men and women.

B. Further experimental details

B.1. Messages sent

In this section, we provide the English translation of the messages sent in the experiment.

Baseline ($n = 3768$: 1585 women and 2183 men): No message was sent.

Reminder ($n = 2823$: 1165 women and 1658 men): Dear XXX: A kind reminder that you have a repayment due on YEAR-MONTH-DATE, please pay back on time. For more inquiries, call SERVICE NUMBER.

Norm ($n = 2807$: 1166 women and 1641 men): Dear XXX: A kind reminder that you have a repayment due on YEAR-MONTH-DATE, please pay back on time. We trust you to repay your loan on time as about 95% of borrowers have successfully repaid. For more inquiries, call SERVICE NUMBER.

Shame ($n = 2789$: 1161 women and 1628 men):

(The message sent to the borrower) Dear XXX: A kind reminder that you have a repayment due on YEAR-MONTH-DATE, please pay back on time. If you do not repay your loan on time, we will inform the endorsers you nominated. For more inquiries, call SERVICE NUMBER.

(The message sent to the endorsers) Dear XXX: A kind reminder that XXX, whom you endorsed, has not repaid the loan due on YEAR-MONTH-DATE, please inform him/her to repay as soon as possible to avoid credit loss. For more inquiries, call SERVICE NUMBER.

Reward ($n = 2815$: 1172 women and 1643 men): Dear XXX: A kind reminder that you have a repayment due on YEAR-MONTH-DATE, please pay back on time. If you repay your loan on time, we will give you a five-percent interest rate discount for your future loans. For more inquiries, call SERVICE NUMBER.

Punishment ($n = 2543$: 1022 women and 1521 men): Dear XXX: A kind reminder that you have a repayment due on YEAR-MONTH-DATE, please pay back on time. If you do not repay your loan on time, we will give you a five-percent interest rate mark-up for your future loans. For more inquiries, call SERVICE NUMBER.

We plot the screenshots of the messages sent to the borrowers and endorsers in the Shame treatment in Figure B.1. The display of the messages for other treatments is similar, and the screenshots are available upon request.

B.2. Balance check

Table B.1 shows the mean values of the demographic variables and loan characteristics for all experimental participants by treatment group. The F-tests suggest that all the demographic and loan information we collected are highly balanced. There is no significant difference in any demographic or loan variables across treatments, which implies that our randomization was successful.

(a) Message sent to borrower

(b) Message sent to endorser

Figure B.1: Screenshot for Shame treatment. This figure shows the original messages sent to both the borrower in Panel (a) and the endorser in Panel (b) for the Shame treatment.

Table B.1: Balance check. This table presents the mean values for demographic, credit, and loan characteristics for each treatment. We also report the F-statistics and p-values testing the null hypothesis that there is no difference in the variable across all treatments.

	Baseline	Reminder	Norm	Shame	Reward	Punishment	F_{stat}	p_{value}
Panel A: Demographic Variables								
Gender (male=1)	0.579	0.587	0.585	0.584	0.584	0.598	0.48	0.80
Age	27.91	27.72	27.85	27.92	27.85	27.92	0.49	0.78
Monthly Income	3896.63	3941.11	3888.19	4007.97	3892.90	3993.33	1.58	0.16
Employment Indicator	0.539	0.537	0.545	0.556	0.562	0.561	1.49	0.19
Higher Education	0.674	0.672	0.684	0.693	0.690	0.684	0.98	0.43
Married Indicator	0.485	0.486	0.501	0.494	0.510	0.483	1.35	0.24
Panel B: Credit Variables								
Credit Score	2.884	2.871	2.865	2.902	2.914	2.894	0.77	0.57
Car Indicator	0.465	0.457	0.470	0.477	0.487	0.460	1.43	0.21
House Indicator	0.616	0.629	0.634	0.631	0.629	0.639	0.84	0.52
Other Loan Indicator	0.239	0.223	0.230	0.236	0.233	0.240	0.62	0.68
Past Borrowing Incidence	2.101	2.084	2.182	2.096	2.043	2.148	1.12	0.35
Historical Loan Amount	7241.39	9025.59	7325.31	8116.39	6652.54	8179.21	2.05	0.07
Panel C: Loan Information Variables								
Loan Amount	5446.39	5518.36	5446.52	5448.30	5430.69	5429.39	0.39	0.86
Loan Term	10.90	10.96	10.93	11.00	11.00	10.90	1.15	0.33
Loan Interest Rate (%)	16.91	17.04	16.95	16.92	16.97	16.97	0.65	0.66
Family Endorsers	2.583	2.669	2.608	2.570	2.580	2.577	1.82	0.10

C. Further analysis

C.1. Days overdue

Figure C.1 plots the average number of days a loan is overdue across treatments and by gender. We record 0 overdue day for borrowers who repaid on time and truncate the days overdue to 30 where borrowers then go into default and other interventions may occur. For female borrowers, the ranking of the mechanisms from the most effective to the least effective is Norm (0.95 days overdue), Shame (1.12 days overdue), Punishment (2.19 days overdue), Reward (3.09 days overdue), and Reminder (4.67 days overdue). The ranking for male borrowers is Punishment (1.01 days overdue), Reward (1.05 days overdue), Shame (2.15 days overdue), Norm (2.21 days overdue), and Reminder (2.96 days overdue). Women are more sensitive to social incentives, while men are more sensitive to financial incentives. In Table C.1, we regress days overdue on the set of dummy variables specifying treatments and genders as well as other demographic, credit, and loan variables. The coefficients in Column (1) indicate that the results regarding the gender differences are robust to the control variables we considered.

C.2. Default rate

Figure C.2 plots the default rate across treatments and by gender. For women, the ranking of the treatments from the most effective to the least effective is Norm (1.3% default), Shame (2.3% default), Punishment (5.2% default), Reward (5.6% default), and Reminder (8.1% default). For men, the ranking is Punishment (1.6% default), Reward (1.9% default), Shame (3.4% default), Reminder (4.1% default), and Norm (5.0% default). Women are more sensitive to social incentives, while men are more sensitive to financial incentives. We also perform a set of regressions analogous to those in Table 2 to study the robustness of these gender differences, and we report the estimates in Columns (2)–(4) in Table C.1. The results hold true after adding the controls.

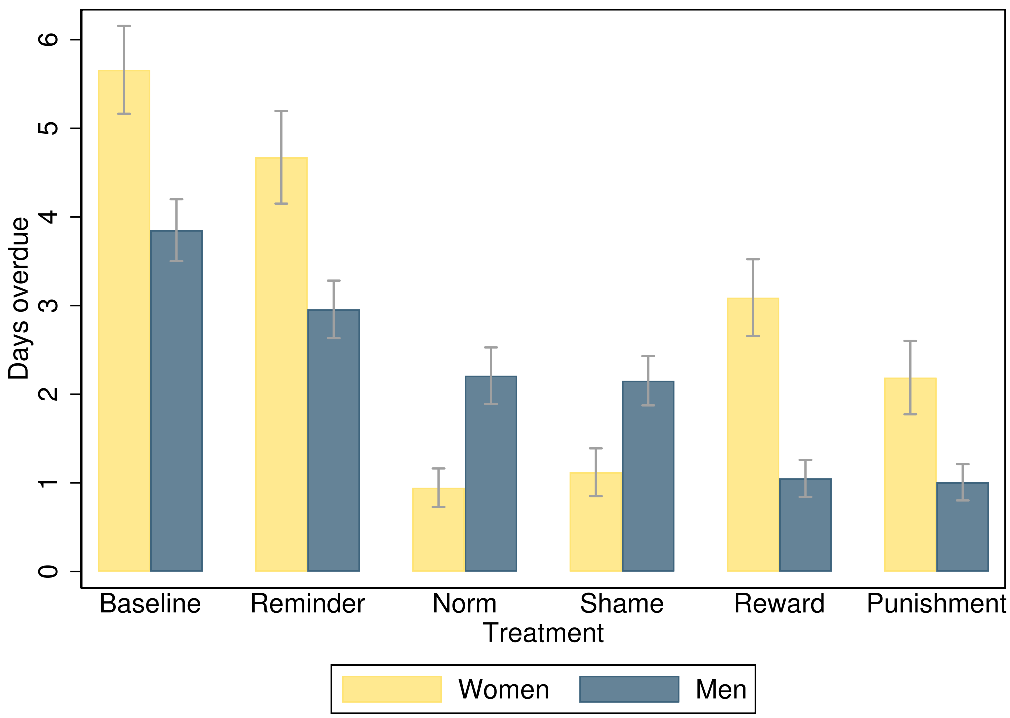


Figure C.1: Days overdue. This figure plots the average overdue days for both genders across treatments. The error bars indicate 95% confidence intervals.

Table C.1: Regressions comparing the intensity of non-compliance across treatments. **This table explores how experimental interventions affect days overdue and default behavior for male and female borrowers. t-statistics are reported in parentheses below each coefficient. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The bottom panel displays the p-values testing the null hypotheses that both genders reduce their overdue rate at the same magnitude in response to the treatment as compared with the baseline.**

	Dependent Variable:			
	Days overdue	Default Indicator		
	(1)	(2)	(3)	(4)
Male baseline	-1.745***	-0.041***	-0.661***	-0.330***
	(-5.81)	(-4.86)	(-4.55)	(-4.41)
Male reminder	-2.635***	-0.052***	-0.974***	-0.473***
	(-9.00)	(-6.16)	(-5.85)	(-5.62)
Male norm	-3.439***	-0.045***	-0.725***	-0.351***
	(-11.79)	(-5.09)	(-4.52)	(-4.29)
Male shame	-3.451***	-0.058***	-1.131***	-0.590***
	(-12.29)	(-7.02)	(-6.64)	(-6.86)
Male reward	-4.496***	-0.071***	-1.760***	-0.851***
	(-16.81)	(-9.11)	(-8.29)	(-8.27)
Male punish	-4.495***	-0.073***	-1.809***	-0.907***
	(-16.75)	(-9.39)	(-7.88)	(-8.39)
Female reminder	-0.984***	-0.018*	-0.268*	-0.147*
	(-2.83)	(-1.78)	(-1.77)	(-1.87)
Female norm	-4.450***	-0.075***	-2.135***	-1.022***
	(-16.70)	(-9.83)	(-7.58)	(-7.79)
Female shame	-4.285***	-0.065***	-1.493***	-0.729***
	(-15.51)	(-7.99)	(-6.76)	(-6.72)
Female reward	-2.389***	-0.034***	-0.573***	-0.288***
	(-7.43)	(-3.65)	(-3.43)	(-3.38)
Female punish	-3.389***	-0.042***	-0.654***	-0.340***
	(-10.77)	(-4.40)	(-3.73)	(-3.76)
Controls	Yes	Yes	Yes	Yes
Method	OLS	OLS	Logit	Probit
N	17,545	17,545	17,545	17,545
DID reminder	p =0.824	p =0.580	p =0.840	p =0.977
DID norm	p =0.000	p =0.000	p =0.000	p =0.000
DID shame	p =0.000	p =0.000	p =0.000	p =0.001
DID reward	p =0.034	p =0.073	p =0.051	p =0.079
DID punish	p =0.088	p =0.390	p =0.084	p =0.087

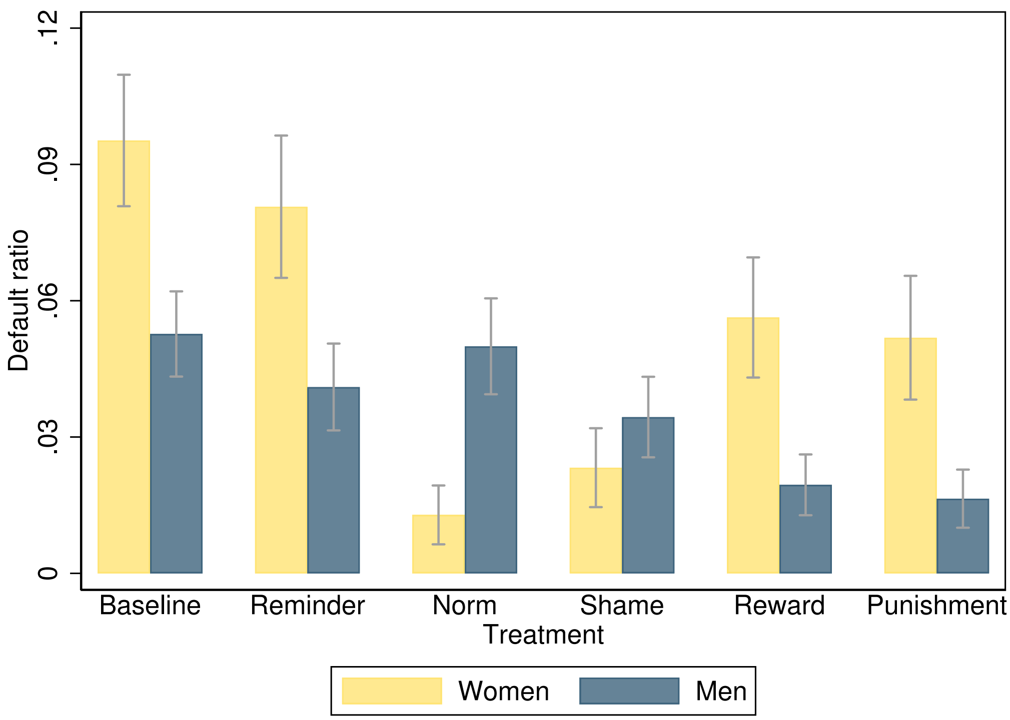


Figure C.2: Default ratio. This figure plots the average default ratio across treatments by gender. The error bars plot the 95% confidence interval.