Department of Economics

ISSN number 1441-5429

Discussion number 08/20

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

July 29, 2020

Difang Huang
Monash University

Zhengyang Bao
Monash University

Abstract

We followed 58,345 borrowers from a peer-to-peer lending platform to study how females and males react to enforcement mechanisms differently. In the experiment, borrowers were randomized into treatments where they received different text messages urging for timely repayment if they had loans due the “next day”. Compared to a reminder message, the messages inducing social pressures and financial incentives reduced the overdue rate for both genders. However, females were more responsive to messages producing social pressures, while males were more responsive to financial incentives. The results imply the potential importance of a gender-dependent mechanism to enhance compliance.

Keywords: Gender differences; Natural field experiment; Enforcement mechanism

JEL codes: C93, D91, J16

The authors gratefully acknowledge the comments from Lata Gangadharan and Andreas Leibbrandt.

@2020 Difang Huang and Zhengyang Bao.

All rights reserved. No part of this paper may be reproduced in any form, or stored in a retrieval system, without the prior written permission of the author.
monash.edu/business/economics

ABN 12 377 614 012

CRICOS Provider No. 00008C

1 Introduction

Females and males have some fundamental differences in preferences and behaviors. For example, females tend to be more risk-averse, less prone to competitive environments, and respond to financial incentives differently from males. These gender differences hint that males and females may react differently to some enforcement mechanisms. 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 on borrowers from a large peer-to-peer (P2P) lending platform in China.¹

The P2P lending platform collects funds from the lenders and makes loans to the borrowers, similar to a financial intermediary. It attracted 6.54 million users 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 testbed for gender differences in responding 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 nation ID, which contains the gender of the person at birth. Second, the non-compliance (overdue) rate is relatively gender-neutral (male borrowers contribute to about 54% of the late repayments), so the results are less likely to suffer from biases caused by selection or stereotype. Third, it has a large sample size with a moderate overdue 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 errors. Fifth, the project may find a useful gender-specific mechanism to deter late repayment, which is meaningful economically and socially.

In the experiment, male and female borrowers with no overdue history are randomized into one of the treatments and receive text messages from the platform asking them to make

¹The popularity of P2P lending is not restricted to China. [de Roure, Pelizzon, and Tasca \(2016\)](#) provide empirical evidence that the P2P lending in Germany substitutes traditional banking for the role of credit access of high-risk borrowers, [Di Maggio and Yao \(2018\)](#) show that in the US, compared with conventional banking, P2P lending enables borrowers with immediate consumption needs to finance their expenses beyond capability.

the repayment on time. The content of the message varies across treatments. To encourage the repayment, we embed into the messages one of three types of enforcement mechanisms that are widely explored in the literature. The first mechanism is a simple reminder message to mitigate the delays caused by limited attention. The second mechanism invokes social pressure. In particular, we test how borrowers react to information about the social norm (i.e., most borrowers repaid on time) and a threat to inform the endorsers (usually people close to the borrower) if the borrower fails to repay on time. The third mechanism introduces financial incentives. We test rewards for compliance (i.e., the platform offered to decrease the interest rate for future borrowing if the borrower made the repayment on time) and punishments for non-compliance (i.e., the platform threatened to increase the interest rate if the borrower failed to make the repayment on time).

We find that social pressures and financial incentives successfully reduce the overdue rate for both male and female borrowers as compared with the scenario 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 enforcement mechanism for females to reduce the probability of overdue is the social pressures from notifying the endorsers, while the most effective one for their male counterparts is a 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, we can further cut the overdue rate to 13.7%— a 20% improvement. These results imply the potential merit for a gender-dependent enforcement mechanism to deter non-compliance.

This project contributes to the literature studying gender differences. Ample evidence finds male and females differ in time and risk preference (Castillo, Leo, and Petrie, 2019; Charness and Gneezy, 2012; Dittrich and Leipold, 2014; Dohmen et al., 2011; Eckel and Grossman, 2002, 2008; Xie, Page, and Hardy, 2017), social preference (Andreoni and Vesterlund, 2001; Beugnot et al., 2019; Castillo et al., 2013; Chaudhuri and Gangadharan, 2002; Eckel and Grossman, 1996, 2001; Friesen and Gangadharan, 2012), and competitiveness (Datta Gupta, Poulsen, and Villeval, 2013; Flory et al., 2018; Gneezy and Rustichini, 2004; Gneezy, Niederle, and Rustichini, 2003; Niederle and Vesterlund, 2007; Wozniak,

Harbaugh, and Mayr, 2014).² These differences are likely contributors to gender gaps in labor and financial markets outcomes (Blau and Kahn, 2000; Dwyer, Gilkeson, and List, 2002; Eckel and Fullbrunn, 2015; Tate and Yang, 2015). A significant proportion of the existing literature focuses on the origin of these differences (Gneezy, Leonard, and List, 2006; Hoffman, Gneezy, and List, 2011) and promoting gender equality in the market outcomes (Bao and Leibbrandt, 2020; Erkal, Gangadharan, and Xiao, 2019; Flory, Leibbrandt, and List, 2015; Leibbrandt, Wang, and Foo, 2018; Niederle, Segal, and Vesterlund, 2013). Contrary to diminishing the gender gap, we utilize the gender differences in the responsiveness to different enforcement mechanisms and propose a gender-dependent mechanism to improve compliance with financial commitments.

We also contribute to the literature exploring enforcement mechanisms, such as using reminders (Du et al., 2020), social preferences (Brusa, Luo, and Fang, 2019; Hage, Söderholm, and Berglund, 2009; Kahan and Posner, 1999; Kandori, 1992), and financial incentives (Andreoni, Harbaugh, and Vesterlund, 2003; Luo et al., 2017; Xiao and Houser, 2005); to deter non-compliance activities. The majority of the studies focus on a single mechanism (Abbink et al., 2014; Bhanot, 2017; Bao, Gangadharan, and Leister, 2020; Chen and Rey, 2013; Xiao and Houser, 2011), while we compare the effectiveness of several different enforcement mechanisms across genders and find the strength of customizing enforcement mechanisms according to the characteristic (i.e., gender) of the potential non-compliant.

²Interested readers may find Croson and Gneezy (2009) for a comprehensive review of gender differences in preferences.

2 Experimental design

2.1 The P2P lending platform

P2P lending is an important channel to provide credit in China because many households have limited access to the formal banking system.³ The platform where the experiment was conducted was a large size P2P lending platform in China.

Before borrowing money, a borrower has to register with the platform with his/her personal information, including a copy of the national ID, which contains information about his/her gender, age, and birthplace. The borrower also needs to submit five friends and/or family members as endorsers who may be contacted if the borrower fails to pay the money back.

To borrow money, the borrower has to make an application specifying the amount of money requested and the loan duration (term). The platform evaluates the application and chooses to either reject the application or negotiate the interest rate with the borrower. The borrower gets the loan if they agree on a mutually acceptable interest rate. 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, and no partial payment is allowed.⁴ Each borrower can only have one outstanding loan at a time, which implies that she/he can not borrow a new fund from the platform to pay back the existing one.

If repayment is overdue, the borrower is given a 29-day window to repay the loan with a daily penalty, which is much higher than the corresponding interest rate. If the borrower still fails to pay after the 29-day window, the loan is considered a default, and the legal process to collect the debt starts.

³See [Jiang et al. \(2018\)](#) and [Liao et al. \(2018\)](#) for the documentation on the distinctive features of the peer-to-peer industry in China.

⁴The annualized interest rate ranges from 7% to 24% and the duration of the loan the borrower can choose has the minimum of one month and the maximum of twenty-four months.

2.2 The treatments

Limited attention and moral hazard were two likely contributors to not paying back on time. To mitigate forgetfulness, we designed a reminder treatment (Reminder thereafter) which sent a reminder text to the borrower one day before the repayment was overdue. There were several ways to reduce moral hazard. We considered using social pressure and monetary incentives in the experiment. To induce social pressure, we sent the borrower a message stating that most borrowers paid back on time in the social norm treatment (Norm thereafter); and a message threatening to contact the endorsers the borrower specified if the borrower failed to repay on time in the shame treatment (Shame thereafter). To create monetary incentives without changing the budget conditions in the current period, we manipulated the interest rate for future borrowing. In particular, we sent a message to the borrower guaranteed to reduce the interest rate by 5% for future borrowing if she/he repaid on time in the reward treatment (Reward thereafter); and threatened to increase the interest by 5% if the repayment was overdue in the punishment treatment (Punishment thereafter). The platform implemented all the incentives promised in the messages, so there was no deception in the experiment. To conclude, we had the following treatments (the exact translation of the messages is provided in the Appendix):

Baseline ($n = 3768$: 1585 females and 2183 males): no message was sent.

Reminder ($n = 2823$: 1165 females and 1658 males): a simple reminder message was sent asking the participant to repay on time.

Norm ($n = 2807$: 1166 females and 1641 males): a message stated that most borrowers made their repayment on time and asked the participant to do the same as.

Shame ($n = 2789$: 1161 females and 1628 males): a message stated that her/his endorsers would be notified if the participant did not make the repayment on time.

Reward ($n = 2815$: 1172 females and 1643 males): a message to reduce the interest rate for the future loan if the participant made the repayment on time.

Punishment ($n = 2543$: 1022 females and 1521 males): a message to increase the interest rate for the future loan if the participant made the repayment on time.

2.3 The procedures

The experiment was conducted between January 2017 and March 2017. During which, borrowers with no overdue record were potential participants for our experiment and they could only participate in one of the treatments at most. We excluded borrowers with overdue history because they had received other messages from the platform that can contaminate the incentives created by this experiment. Any borrower passed this screening process enters our subject pool if the person had not made the repayment at least one day before the due date. In total, we followed 58,345 qualified borrowers and 17,545 became the participants of the experiment.⁵ The system randomly assigned each participant into one of the treatments. Each participant was notified with a message, which varied across treatments, except the borrowers in the baseline did not receive any message. To avoid complications, the platform did not contact the participants again until the repayment was 29 days overdue if they had not repaid earlier.

2.4 Hypotheses

Standard theory assumes that people maximize their monetary outcomes without making mistakes. Therefore, without financial incentives, the theory predicts that none of Reminder, Norm, or Shame can encourage people to stick to financial commitments. However, in the real-world, people suffer from limited attention and have other-regarding preferences (see [Bursztyn et al., 2019](#); [Cappelen et al., 2020](#), for example). Therefore, a reminder

⁵The 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 orthogonal to the characteristics of the participants, we pooled these participants into the baseline. The characteristics of the participants across treatments were highly balanced. Therefore, the failure did not invalidate our conclusions. Balance checks are reported in the Appendix.

message may reduce late repayment caused by forgetfulness, and messages that induce social pressures may ease the severity of the moral hazard. Thus, with the standard and behavioral theories, we expect all our interventions to reduce the overdue rate.

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

So far, we do not take gender differences into account. Research from psychology and neural sciences suggests that women are more sensitive to social cues in determining appropriate behavior (Chen et al., 2019; Gilligan, 1993). Therefore, we hypothesize that females respond more than males to Norm. In Shame, the effectiveness of the message also depends on the endorsers the borrower has chosen. From the pre-experiment data provided by the platform, female borrowers nominate 3.15 family members (1.85 friends) on average while the number for male borrowers is only 2.2 family members (2.8 friends). In Chinese culture, a kinship is considered to be much stronger than a friendship. Thus, the Shame from a family member is likely to carry more weight than that from a friend. As a result, we expect Shame to be more effective for female borrowers as well.

Hypothesis 2 *Females respond more to the social incentives.*

For financial incentives, several studies reveal that males are more sensitive to monetary losses and benefits (e.g., Kulich et al., 2011; Pokorny, 2008), despite that Bandiera et al. (2016) argues that such gender difference is small and study-specific.

Loss aversion (i.e., losses weigh more heavily than gains in decision processes) is another behavioral trait that may affect male and female borrowers differently. Since Kahneman and Tversky (1979), ample evidence finds that loss aversion is prevalent (e.g., Abdellaoui, Bleichrodt, and Paraschiv, 2007; Andersson et al., 2016; Imas, Sadoff, and Samek, 2017). Therefore, we expect Punishment to be more effective than Reward to make both genders pay back on time. In addition, some studies find that males are more susceptible to loss aversion (Grolleau, Kocher, and Sutan, 2016; Schmidt and Traub, 2002). Based on the above evidence, we conjecture that females respond less than males to the financial incentives.

Hypothesis 3 *Males respond more to the financial incentives.*

3 Results

We use the overdue rate as a measure of non-compliance to study how the treatment interventions affect behavior. We use the overdue rate as the outcome variable is because all of our treatments create incentives based on the repayment due date. We report the results based on other measures of non-compliance such as the duration of overdue and the default rate in the Appendix. The main findings are robust to different measures of non-compliance in the sense that female borrowers are more responsive to social incentives, while male borrowers are more sensitive to financial incentives.

3.1 Overdue rate in each treatment

Figure 1 plots the overdue rate across treatments. The messages significantly reduced the overdue rate from 37.1% in the Baseline to 34.2% in Reminder ($p = 0.0167$), 24.4% in Norm ($p < 0.0001$), 24.1% in Shame ($p < 0.0001$), 21.1% in Reward ($p < 0.0001$) and 17.0% in Punishment ($p < 0.0001$).⁶ The data confirm Hypothesis 1 that participants respond to both financial and non-financial incentives. The data also shows that borrowers are more responsive to financial incentives (i.e., Reward+Punishment) than social incentives (i.e., Norm+Shame), $p < 0.0001$. In summary, Punishment is the most effective deterrence mechanism in the experiment if we do not know the genders of the participants.

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

⁶Without further specification, all p -values reported in the paper come from two-sided t -tests.

Figure 1: Overdue ratio

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

3.2 Gender differences in the overdue rate

Figure 2 decomposes the rate by gender to study if there are gender differences in the reaction to the messages. As conjectured in Hypothesis 2 and 3, the responsiveness to interventions is not gender-neutral. For females, the ranking of the mechanisms from the most effective to the least effective is Shame (14.0% overdue), Norm (19.2% overdue), Punishment (22.2% overdue), Reward (27.2% overdue), Reminder (37.2% overdue); while the ranking for the males is Punishment (13.5% overdue), Reward (16.7% overdue), Norm (28.1% overdue), Shame (31.3% overdue), Reminder (32.1% overdue). 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%.

Figure 2: Overdue ratio by genders

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

3.3 Gender differences in the responsiveness to mechanisms

Figure 3 provides more fine-grained comparisons in the effectiveness of the enforcement mechanisms between females and males. There is no significant difference between female and male borrowers in responding to reminder messages (male: -3.4% v.s. female: -2.4%, $p = 0.6236$). However, there are considerable gender differences in other treatments. In Norm and Shame, females reduced the overdue rate by 21.5% and 26.7% in absolute terms, respectively, while the corresponding numbers for males were 6.3% and 3.1% ($p < 0.0001$ for both comparisons). In Reward and Punishment, females reduced the overdue rate by 13.5% and 18.5% in absolute terms, respectively, while the corresponding numbers for males were 17.7% and 20.9% ($p = 0.054$ for Reward and $p = 0.2869$ for Punishment).

Figure 3: Change in the overdue ratio by genders

Notes: This figure plots the changes in the overdue ratio as compared to the Baseline, in absolute values, across treatments and by genders. The error bars plot the 95% confidence interval.

To test the robustness of the above results, we regress the overdue rate on a set of dummy variables specifying treatments and genders plus a set of control variables including age (in years), the natural logarithm of self-reported income, the interest of the loan, the duration of the loan and the credit score that the P2P platform assigned to each participant on a scale of 1 to 7 with 1 being the most credible. Table 1 shows the estimation results using the linear probability model, logit model, and probit model in column (1), (2), (3), respectively. The results from Figure 2 and 3 are robust to the control variables and estimation methods we considered. For the control variables we included, only the credit score significantly predicts the overdue ratio, and none of age, income, interest rate, and duration of the loan can explain the overdue rate.

Result 2: Females respond more to the social incentives. The most effective mechanism to deter the breach of the repayment commitment for females is Shame.

Result 3: Males respond more to the financial incentives. The most effective mechanism to deter the breach of the repayment commitment for males is Punishment.

4 Discussion

We study how male and female borrowers respond to different messages urging for timely repayment in a P2P context. While we find all of the interventions studied in the experiment successfully deter non-compliance to the repayment commitments as expected, there are significant gender differences. If the P2P platform does not want to offer financial incentives as in Punishment and Reward, Shame works the best for females while Norm works the best for males.⁷ With the financial incentives, Shame still works the best for females while Punishment works the best for males. These results imply that the P2P platform can customize the enforcement mechanisms, among the available options, based on the information about the borrower to improve the compliance.

As some studies have pointed out, gender differences sometimes are context and culture-specific (Andersen et al., 2018; Houser and Schunk, 2009; Wang, Houser, and Xu, 2018). Whether these results hold in a larger scale affects the optimal practice for the P2P lending industry (Al-Ubaydli, List, and Suskind, 2019; Al-Ubaydli and List, 2013). We discuss the external validity of this study using the conditions proposed by List (2020). First, our sample is highly representative of borrowers from the P2P lending platform. All participants are randomly selected. 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 decisions in a natural environment. As a result, we expect our results to be valid for a larger sample size and borrower populations from similar P2P lending platforms.

However, caution should be exerted to apply the same interventions to other contexts as the characteristics of the population can be very different. Nevertheless, our experiment provides a methodological tool to improve the effectiveness of enforcement mechanisms and suggest the enforcers to consider experimenting with the available enforcement mechanisms,

⁷The P2P platform may choose not to use Punishment and Reward as enforcement tools. Increasing the interest rate is impractical if the rate is already at the upper bound of the legal rate in China while decreasing the interest rate can reduce the platform's revenue.

before implement on a large scale, to find the optimal way to customize the messages according to the characteristics of potential violators to achieve the best outcome.

Table 1: Regressions comparing overdue rates across treatments

	DV: Overdue Rate		
	(1)	(2)	(3)
Male baseline	-0.061*** (0.016)	-0.258*** (0.070)	-0.159*** (0.043)
Male reminder	-0.089*** (0.017)	-0.384*** (0.076)	-0.236*** (0.047)
Male norm	-0.128*** (0.017)	-0.574*** (0.078)	-0.351*** (0.047)
Male shame	-0.096*** (0.017)	-0.418*** (0.077)	-0.257*** (0.047)
Male reward	-0.242*** (0.016)	-1.238*** (0.086)	-0.735*** (0.050)
Male punish	-0.265*** (0.016)	-1.445*** (0.093)	-0.846*** (0.053)
Female reminder	-0.039** (0.019)	-0.167** (0.080)	-0.103** (0.050)
Female norm	-0.220*** (0.017)	-1.083*** (0.091)	-0.648*** (0.053)
Female shame	-0.272*** (0.016)	-1.465*** (0.100)	-0.861*** (0.056)
Female reward	-0.139*** (0.018)	-0.629*** (0.084)	-0.384*** (0.051)
Female punish	-0.178*** (0.018)	-0.845*** (0.093)	-0.510*** (0.055)
Age	0.000 (0.001)	0.002 (0.003)	0.001 (0.002)
Ln(Income)	0.000 (0.007)	0.000 (0.035)	-0.001 (0.021)
Interest rate (%)	0.000 (0.001)	0.001 (0.008)	0.001 (0.005)
Loan terms (months)	-0.002 (0.001)	-0.010 (0.008)	-0.006 (0.005)
Credit score	0.006* (0.003)	0.029* (0.017)	0.017* (0.010)
Constant	0.398*** (0.063)	-0.432 (0.326)	-0.263 (0.194)
Estimation method	OLS	Logit	Probit

Notes: This table reports OLS, logit and probit estimates of overdue rate under different treatments across genders. We use 17,545 observations for each regression. The standard errors are reported in brackets below each coefficient. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

References

- Abbink, K., U. Dasgupta, L. Gangadharan, and T. Jain. 2014. Letting the briber go free: An experiment on mitigating harassment bribes. *Journal of Public Economics* 111:17–28.
- Abdellaoui, M., H. Bleichrodt, and C. Paraschiv. 2007. Loss aversion under prospect theory: A parameter-free measurement. *Management Science* 53:1659–74.
- Al-Ubaydli, O., and J. A. List. 2013. On the generalizability of experimental results in economics. *Methods of Modern Experimental Economics*.
- Al-Ubaydli, O., J. A. List, and D. Suskind. 2019. The science of using science: Towards an understanding of the threats to scaling experiments. Working paper.
- Andersen, S., S. Ertac, U. Gneezy, J. A. List, and S. Maximiano. 2018. On the cultural basis of gender differences in negotiation. *Experimental Economics* 21:757–78.
- Andersson, O., H. J. Holm, J.-R. Tyran, and E. Wengström. 2016. Deciding for others reduces loss aversion. *Management Science* 62:29–36.
- Andreoni, J., W. Harbaugh, and L. Vesterlund. 2003. The carrot or the stick: Rewards, punishments, and cooperation. *American Economic Review* 93:893–902.
- Andreoni, J., and L. Vesterlund. 2001. Which is the fair sex? gender differences in altruism. *The Quarterly Journal of Economics* 116:293–312.
- Bandiera, O., G. Fischer, A. Prat, and E. Ytsma. 2016. Do women respond less to performance pay? building evidence from multiple experiments. Working paper.
- Bao, Z., L. Gangadharan, and M. Leister. 2020. Deterrence using peer information. Working paper.
- Bao, Z., and A. Leibbrandt. 2020. Tournaments with safeguards: A blessing or a curse for women? Working paper.

- Beugnot, J., B. Fortin, G. Lacroix, and M. C. Villeval. 2019. Gender and peer effects on performance in social networks. *European Economic Review* 113:207–24.
- Bhanot, S. P. 2017. Rank and response: A field experiment on peer information and water use behavior. *Journal of Economic Psychology* 62:155–72.
- Blau, F. D., and L. M. Kahn. 2000. Gender differences in pay. *Journal of Economic Perspectives* 14:75–99.
- Brusa, F., X. Luo, and Z. Fang. 2019. The power of non-monetary incentive: Experimental evidence from p2p lending in china. Working paper.
- Bursztyn, L., S. Fiorin, D. Gottlieb, and M. Kanz. 2019. Moral incentives in credit card debt repayment: evidence from a field experiment. *Journal of Political Economy* 127:1641–83.
- Cappelen, A., J. List, A. Samek, and B. Tungodden. 2020. The effect of early-childhood education on social preferences. *Journal of Political Economy* 128:000–.
- Castillo, M., G. Leo, and R. Petrie. 2019. Room composition effects on risk taking by gender. *Experimental Economics* 1–17.
- Castillo, M., R. Petrie, M. Torero, and L. Vesterlund. 2013. Gender differences in bargaining outcomes: A field experiment on discrimination. *Journal of Public Economics* 99:35–48.
- Charness, G., and U. Gneezy. 2012. Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization* 83:50–8.
- Chaudhuri, A., and L. Gangadharan. 2002. Gender differences in trust and reciprocity. Working paper.
- Chen, W., S. Zhang, O. Turel, Y. Peng, H. Chen, and Q. He. 2019. Sex-based differences in right dorsolateral prefrontal cortex roles in fairness norm compliance. *Behavioural Brain Research* 361:104–12.

- Chen, Z., and P. Rey. 2013. On the design of leniency programs. *The Journal of Law and Economics* 56:917–57.
- Croson, R., and U. Gneezy. 2009. Gender differences in preferences. *Journal of Economic Literature* 47:448–74.
- Datta Gupta, N., A. Poulsen, and M. C. Villeval. 2013. Gender matching and competitiveness: Experimental evidence. *Economic Inquiry* 51:816–35.
- de Roure, C., L. Pelizzon, and P. Tasca. 2016. How does p2p lending fit into the consumer credit market? Working paper.
- Di Maggio, M., and V. W. Yao. 2018. Fintech borrowers: lax-screening or cream-skimming? Harvard Business School and Georgia State University Working Paper.
- Dittrich, M., and K. Leipold. 2014. Gender differences in time preferences. *Economics Letters* 122:413–5.
- Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner. 2011. Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association* 9:522–50.
- Du, N., L. Li, T. Lu, and X. Lu. 2020. Prosocial compliance in p2p lending: A natural field experiment. *Management Science* 66:315–33.
- Dwyer, P. D., J. H. Gilkeson, and J. A. List. 2002. Gender differences in revealed risk taking: evidence from mutual fund investors. *Economics Letters* 76:151–8.
- Eckel, C. C., and S. C. Fullbrunn. 2015. Thar she blows? gender, competition, and bubbles in experimental asset markets. *American Economic Review* 105:906–20.
- Eckel, C. C., and P. J. Grossman. 1996. Altruism in anonymous dictator games. *Games and Economic Behavior* 16:181–91.

- . 2001. Chivalry and solidarity in ultimatum games. *Economic Inquiry* 39:171–88.
- . 2002. Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior* 23:281–95.
- . 2008. Men, women and risk aversion: Experimental evidence. *Handbook of Experimental Economics Results* 1:1061–73.
- Erkal, N., L. Gangadharan, and E. Xiao. 2019. Leadership selection: Can changing the default break the glass ceiling? Working paper.
- Flory, J. A., U. Gneezy, K. L. Leonard, and J. A. List. 2018. Gender, age, and competition: A disappearing gap? *Journal of Economic Behavior & Organization* 150:256–76.
- Flory, J. A., A. Leibbrandt, and J. A. List. 2015. Do competitive workplaces deter female workers? a large-scale natural field experiment on job entry decisions. *The Review of Economic Studies* 82:122–55.
- Friesen, L., and L. Gangadharan. 2012. Individual level evidence of dishonesty and the gender effect. *Economics Letters* 117:624–6.
- Gilligan, C. 1993. *In a different voice psychological theory and women's development*. Harvard University Press.
- Gneezy, U., K. L. Leonard, and J. A. List. 2006. Gender differences in competition: the role of socialization. Working paper.
- Gneezy, U., M. Niederle, and A. Rustichini. 2003. Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics* 118:1049–74.
- Gneezy, U., and A. Rustichini. 2004. Gender and competition at a young age. *American Economic Review* 94:377–81.

- Grolleau, G., M. G. Kocher, and A. Sutan. 2016. Cheating and loss aversion: Do people cheat more to avoid a loss? *Management Science* 62:3428–38.
- Hage, O., P. Söderholm, and C. Berglund. 2009. Norms and economic motivation in household recycling: Empirical evidence from sweden. *Resources, Conservation and Recycling* 53:155–65.
- Hoffman, M., U. Gneezy, and J. A. List. 2011. Nurture affects gender differences in spatial abilities. *Proceedings of the National Academy of Sciences* 108:14786–8.
- Houser, D., and D. Schunk. 2009. Social environments with competitive pressure: Gender effects in the decisions of german schoolchildren. *Journal of Economic Psychology* 30:634–41.
- Imas, A., S. Sadoff, and A. Samek. 2017. Do people anticipate loss aversion? *Management Science* 63:1271–84.
- Jiang, J., L. Liao, Z. Wang, and X. Zhang. 2018. Government affiliation and fintech industry: The peer-to-peer lending platforms in china. Working paper.
- Kahan, D. M., and E. A. Posner. 1999. Shaming white-collar criminals: A proposal for reform of the federal sentencing guidelines. *The Journal of Law and Economics* 42:365–92.
- Kahneman, D., and A. Tversky. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–91.
- Kandori, M. 1992. Social norms and community enforcement. *The Review of Economic Studies* 59:63–80.
- Kulich, C., G. Trojanowski, M. K. Ryan, S. Alexander Haslam, and L. D. Renneboog. 2011. Who gets the carrot and who gets the stick? evidence of gender disparities in executive remuneration. *Strategic Management Journal* 32:301–21.

- Leibbrandt, A., L. C. Wang, and C. Foo. 2018. Gender quotas, competitions, and peer review: Experimental evidence on the backlash against women. *Management Science* 64:3501–16.
- Liao, L., X. Martin, N. Wang, Z. Wang, and J. Yang. 2018. Carrot or stick? evidence from a pair of randomized field experiments testing lender information sharing hypotheses. Working paper.
- List, J. A. 2020. Non est disputandum de generalizability? a glimpse into the external validity trial. Working paper.
- Luo, X., K. Wittkowski, Z. Fang, and J. Aspara. 2017. Consumer debt collection with social pressure and financial incentives: field experiments. Working paper.
- Niederle, M., C. Segal, and L. Vesterlund. 2013. How costly is diversity? affirmative action in light of gender differences in competitiveness. *Management Science* 59:1–16.
- Niederle, M., and L. Vesterlund. 2007. Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics* 122:1067–101.
- Pokorny, K. 2008. Pay—but do not pay too much: An experimental study on the impact of incentives. *Journal of Economic Behavior & Organization* 66:251–64.
- Schmidt, U., and S. Traub. 2002. An experimental test of loss aversion. *Journal of risk and Uncertainty* 25:233–49.
- Tate, G., and L. Yang. 2015. The bright side of corporate diversification: Evidence from internal labor markets. *The Review of Financial Studies* 28:2203–49.
- Wang, J., D. Houser, and H. Xu. 2018. Culture, gender and asset prices: Experimental evidence from the us and china. *Journal of Economic Behavior & Organization* 155:253–87.

Wozniak, D., W. T. Harbaugh, and U. Mayr. 2014. The menstrual cycle and performance feedback alter gender differences in competitive choices. *Journal of Labor Economics* 32:161–98.

Xiao, E., and D. Houser. 2005. Emotion expression in human punishment behavior. *Proceedings of the National Academy of Sciences* 102:7398–401.

———. 2011. Punish in public. *Journal of Public Economics* 95:1006–17.

Xie, Z., L. Page, and B. Hardy. 2017. Investigating gender differences under time pressure in financial risk taking. *Frontiers in Behavioral Neuroscience* 11:246–.

Appendix A

Messages sent

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

Baseline ($n = 3768$: 1585 females and 2183 males): no message was sent.

Reminder ($n = 2823$: 1165 females and 1658 males): Dear XXX: A kind reminder that you have a repayment due at YEAR-MONTH-DATE, please payback on time. For more inquires call SERVICE NUMBER.

Social norm ($n = 2807$: 1166 females and 1641 males): Dear XXX: A kind reminder that you have a repayment due at YEAR-MONTH-DATE, please payback on time. We trust you to repay your loan on time as about 95% of borrowers have successfully repaid. For more inquires call SERVICE NUMBER.

Shame ($n = 2789$: 1161 females and 1628 males):

(the message sent to the borrower) Dear XXX: A kind reminder that you have a repayment due at YEAR-MONTH-DATE, please payback on time. If you do not repay your loan on time, we will inform the endorsers you nominated. For more inquires 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 at YEAR-MONTH-DATE, please inform him/her to repay as soon as possible to avoid credit loss. For more inquires call SERVICE NUMBER.

Reward ($n = 2815$: 1172 females and 1643 males): Dear XXX: A kind reminder that you have a repayment due at YEAR-MONTH-DATE, please payback on time. If you repay your loan on time, we will give you a five-percent interest rate discount for your next loan. For more inquires call SERVICE NUMBER.

Punishment ($n = 2543$: 1022 females and 1521 males): Dear XXX: A kind reminder that you have a repayment due at YEAR-MONTH-DATE, please payback on time. If you

do not repay your loan on time, we will give you a five-percent interest rate mark-up for your next loan. For more inquires call SERVICE NUMBER.

We plot the screenshots of the messages sent to the borrowers and endorsers in the Shame treatment in Figure 4. The display of the messages for other treatments is similar and the screenshots are available upon request.

Figure 4: Screenshot for Shame treatment

Notes: This figure shows the original messages sent to both the borrower (the left panel) and endorser (the right panel) for the Shame treatment.

Appendix B

Balance check

Table 1 shows the average value of the demographics of participants and loan characteristics across treatments. The F -tests suggest that all the demographic and loan information we collect are highly balanced. There is no significant difference in any demographic and loan variables across treatments, which implies that our randomization was successful.

Table 1: Balance check

	Demographic characteristics			Loan characteristics		
	Gender (male=1)	Age	Self-reported monthly income	Amount (RMB)	Term (months)	Interest rate (%)
Baseline	0.579	27.91	3896.63	5446.39	10.90	16.91
Reminder	0.587	27.72	3941.11	5518.36	10.96	17.04
Norm	0.585	27.85	3888.19	5446.52	10.93	16.95
Shame	0.584	27.92	4007.97	5448.30	11.00	16.92
Reward	0.584	27.85	3892.90	5430.69	11.00	16.97
Punishment	0.598	27.92	3993.32	5429.39	10.90	16.97
F_{stat}	0.48	0.49	1.58	0.39	1.15	0.65
p_{value}	0.795	0.781	0.163	0.858	0.334	0.658

Notes: This table presents the average demographic and loan characteristics for each treatment. We also report the F -statistic and p -value testing the null hypothesis that there is no difference in the variable across all treatments.

Appendix C

Further analysis

A Days overdue

Figure 5 plots the average overdue days across treatments and by genders. We record 0 overdue day for those repaid on time and truncate the days overdue for the defaulted participants to 30 days as we do not observe the repayment behavior of defaulted participants. The pattern is robust to the choice of the truncation number for defaulted participants. For females, 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), Reminder (4.67 days overdue), while the ranking for the males is Punishment (1.01 days overdue), Reward (1.05 days overdue), Shame (2.15 days overdue), Norm (2.21 days overdue), Reminder (2.96 days overdue). Females are more sensitive to social incentives while males are more sensitive to financial incentives.

B Default rate

Figure 6 plots the default rate across treatments and by genders. For females, the ranking of the mechanisms 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), Reminder (8.1% default), while the ranking for the males is Punishment (1.6% default), Reward (1.9% default), Shame (3.4% default), Reminder (4.1% default), Norm (5.0% default). Females are more sensitive to social incentives while males are more sensitive to financial incentives.

Figure 5: Days overdue

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

Figure 6: Default ratio

Notes: This figure plots the average default ratio across treatments and by genders. The error bars plot the 95% confidence interval.