Norm-Based Enforcement of Promises*

Rebecca Stone† UCLA

Alexander Stremitzer‡ ETH Zurich

May 16, 2019

Abstract

There is ample evidence that people are internally motivated to keep their promises. However, it is unclear whether promises alone create a meaningful level of commitment in many economically relevant situations where the stakes are high. In a between-subject design, we ask subjects to imagine they observed as third parties a promisor breaking her promise and could punish the promisor, at a cost to themselves, for her behavior. Our results suggest that the motivations third parties have to punish promise breakers have the same structure as the moral motivations of those deciding whether or not to keep their promises. That is, the same moral reasons that motivate promisors to keep their promises make third-party observers more likely to punish promise breaking. This suggests that the determinants of promise-keeping behavior will also drive non-legal enforcement mechanisms in relational contract settings and situations where third parties can punish promisors in a decentralized fashion, such that the moral forces of promise keeping can generate commitment even when the stakes are high.

Keywords: promises, norms, first-party enforcement, second-party enforcement, altruistic punishment.

JEL-Classification: K12, L14, D86, D91, C91.

1 Introduction

A series of experimental studies find that in the absence of legal enforcement and reputational concerns promisors are nonetheless motivated to keep their promises even when

*The authors are grateful to Gary Charness and William Hubbard for valuable comments. We are also grateful to seminar audiences at Alabama, ETH Zurich, University of Zurich, and the Annual Conference of the Polish Law and Economics Society. We thank Henry Kim for excellent research assistance.

†UCLA Law School, 385 Charles E. Young Drive, 1242 Law Building, Los Angeles, CA 90095, rebecca.stone@law.ucla.edu (corresponding author).

‡ETH Zurich, Center for Law & Economics, IFW E49, Haldeneggsteig 4, 8092 Zurich, Switzerland, astremitzer@ethz.ch.

they have self-interested reasons to break them (Ellingsen and Johannesson 2004, Charness and Dufwenberg, 2006; Vanberg, 2008; Charness and Dufwenberg, 2010). Promisors are also more likely to keep their promises when a promisee expected the promisor to perform or invested more in reliance on the promise (Dufwenberg and Gneezy, 2000; Charness and Dufwenberg, 2006; Ederer and Stremitzer, 2017; Stone and Stremitzer, 2017). And potential promisors are more sensitive to other's expectations of cooperation and reliance investments when they promised to cooperate than when they merely stated an intention to do so—that is, without making any promises (Mischkowski, Stone, and Stremitzer, 2016. In these ways, promisors seem to be motivated by moral reasons when deciding whether to keep their promises. In particular, they seem to believe that it is morally worse to break a promise the more a promisee has relied upon it.

But the willingness of many to keep their promises in the face of self-interested reason to break them won't guarantee performance of a promise—even one that has been relied upon extensively by the promisee—when the self-interested incentives to break it are large enough or when the promisor is a sophisticated self-interested actor who isn't motivated by moral reasons to keep her promise. Thus, effective second- and third-party enforcement mechanisms are needed to give promisors incentives to perform such promises.

To what extent do the moral forces that seem to underpin promisor behavior influence the behavior of those who can enforce promise keeping? Centralized enforcement systems like the legal system can be, and arguably are, designed to reflect or at least support underlying moral practices of their subjects. But what about decentralized enforcement mechanisms, for example, decentralized enforcement by disinterested third-party observers? If promisor behavior is driven by moral norms that deem it worse to break promises the more they have been relied upon, then we should also expect third-party observers to judge a promisor more harshly when she breaks a promise the more it has been relied upon. And if those third parties have the opportunity to punish such a promisor even at some cost to themselves, then, given theory and evidence that suggests that persons are willing to altruistically enforce prevalent norms, we should expect that those third parties will punish the promisor more harshly the

greater was the promisee's reliance.[1] Reputational concerns that track agents' moral beliefs would then make even self-interested promisors act *as if* they had guilt-averse preferences, thereby reinforcing decentralized first-party mechanisms of enforcement.

In this paper we present experimental evidence in support of this hypothesis. Our subjects report a greater willingness to punish, at some cost to themselves, non-cooperative behavior when a potential promisor made a promise to cooperate than when he instead merely reported an intention to do so holding the potential promisee's reliance constant (promising per se effect). They report a greater willingness to punish non-cooperative behavior the more the other party relied on an expectation of cooperation even when the potential promisor simply stated that he intended to cooperate, while explicitly disclaiming a promise (promising per se effect). Finally, the effect of greater reliance on the third-party's willingness to punish non-cooperative behavior is enhanced when a promise to cooperate was made (interaction effect). That is, promising and reliance shape a third-party's willingness to punish promise-breaking in much the same way that they shape a promisor's willingness to keep her promises in the first place.

Taking these results seriously has an important implication. They suggest that moral reasons to keep promises influence not only promisor behavior but also decentralized enforcement mechanisms. This is in line with Hart & Moore's (2008) argument that contracts set reference points or expectations of contracting parties such that disappointed promisees react by punishing the promisor when these expectations are confounded by "shading" on their performance. Theirs is an argument about second-party enforcement. Our results suggest third-party behavior may similarly be shaped by underlying promissory norms. Our results are also in line with theoretical and empirical evidence that altruistically enforced social norms are central determinants of cooperative behavior (see, e.g. Gintis et al. (2005)).

The remainder of the paper is organized as follows. Section 2 describes the experimental design, our hypotheses and the experimental procedure. Section 3 describes our results.

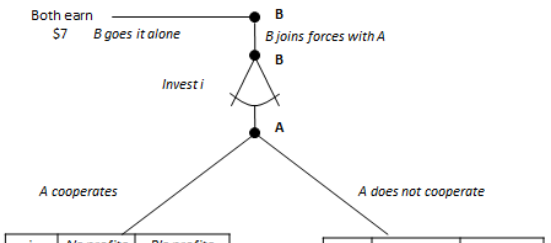
[1] Gintis et al. (2005) provide an overview of the theory and evidence that supports the existence of dispositions towards strong reciprocity. Fehr and Fischbacher (2004) provide experimental evidence of subjects' propensities to punish third parties at a cost to themselves.

Section 4 provides further discussion of those results. Section 5 concludes.

2 Experimental Design & Procedure

2.1 Design and Hypotheses

We built our experiment on the trust game that is depicted in Figure 1. In this game Player B must decide whether or not to join forces with Player A in some venture. Doing so is risky for B because A might exploit B and take the profits from the venture for herself. But if A decides not to exploit B, both players do better than if B had instead pursued the venture on his own. More specifically, the subgame that ensues if B decides to join forces with A is identical to the dictator game that subjects played in our first experiment, with B in the role of the Recipient and A in the role of the Dictator. Both players simply earn profits of 7 if B decides not to join forces with A. Before B makes this decision whether or not to join forces with A, A and B have a conversation in which A attempts to persuade B that she will cooperate with him if he joins forces with her.



i	A's profits	B's profits
0	12	12
1	12	12.25
2	12	12.00
3	12	11.75
4	12	11.50
5	12	11.25
6	12	11.00

i	A's profits	B's profits
0	15	6
1	15	5
2	15	4
3	15	3
4	15	2
5	15	1
6	15	0

Figure 1: Game Witnessed by Third Party.

Subjects were instructed to imagine that they had observed as third parties six permutations of the scenario in which B joined forces with A, but A subsequently decided not to cooperate with B. There were three "Promise conditions" and three "No Promise conditions", each one characterized by a particular investment level chosen by B: 0, 3, and 6.

In the Promise conditions subjects were asked to imagine that A had promised B that she would cooperate with him if he joined forces with her ("I promise that I will cooperate with you"), while in the No Promise conditions they were asked to imagine that A had told B that she planned on cooperating with him without making any promises ("All I can say is that I plan to cooperate with you, though I can't promise that I will do so."). Following the presentation of each scenario, subjects were asked to report the likelihood that they would choose to inflict a punishment on A for not cooperating with B at some small but not insignificant cost to themselves.[2]

Each subject saw six scenarios in a randomized sequence.[3] This allowed us to generate both between- and within-subject data. Subjects' responses to the first condition they saw constitute the between-subject data. Subjects' responses in their entirety constitute the within-subject data.

Our hypotheses follow from our assumption that punishment preferences by third parties mimic promisors' preferences to keep their promises that have been revealed by previous experiments. These experiments suggest that potential promisors are more willing to cooperate when a promise has been made holding constant a potential promisee's expectations of cooperation and reliance on such expectations (a promise per se effect). They suggest that they are more willing to cooperate the greater are promisees' expectations and reliance investments, even when they made no promises to cooperate (an expectation or reliance per se effect). And they suggest that the effect of reliance and expectations on potential promisors' willingness to cooperate are enhanced when they promised to cooperate rather than merely stating their intention to do so (an interaction effect).[4]

[2]Specifically, we told subject that one way of inflicting punishment on A could consist of "boycotting A's business which results in a monetary loss for A." We also reminded them that "[i]nflicting this punishment, however, is not costless to you" as "boycotting a business that you have been buying from forces you to switch to another product requiring you to change your habits, pay more for the other product, consume a product that you like less, etc...)."

[3]The order was not completely random. Subjects either saw all three promise conditions first in a randomized order and then all three no promise conditions in a randomized order or vice versa.

[4]The hypothesis would follow from a model of the observer's preferences that resembles in all important respects the promisor's preferences in Mischkowski, Stone, and Stremitzer (2016). For simplicity, our theoretical model makes the extreme assumption that reliance is irrelevant in the absence of a promise. But our main point is that reliance matters more in the presence of a promise.

Hypothesis 1 *The average reported likelihood of punishment will increase with B's investment level if A made a promise (H1.1) and, possibly, if A made no promise (H1.2). This increase will be greater in the Promise conditions than in the No Promise conditions (H1.3). For all investment levels, the average reported likelihood of punishment will be greater in the Promise conditions than in the No Promise conditions (H1.4).*

2.2 Procedure

We programmed the vignettes using Qualtrics and recruited 1200 subjects from Amazon Mechanical Turk's pool of MTurk workers who had a HIT (Human Intelligence Task) approval rate of 95% or greater. We determined our sample size using a simulation based on pilot data on 300 subjects.[5] Subjects were asked control questions to ascertain whether they had read and understood the scenario, and they were not allowed to proceed until they answered those questions correctly. At the end of the survey subjects were asked several other questions to assess how carefully and honestly they responded to the questions. We also elicited subjects' demographic characteristics.[6]

Before subjects were presented with the scenarios, they were informed that they would be paid \$1.50 for participating, and that the task would take approximately 10 to 15 minutes. The announced hourly wage was therefore \$6 to \$9 per hour. Thus, on average subjects could expect to receive more than the current federal minimum wage (\$7.25 per hour) and expected payments were much higher than the wages MTurk workers typically earn.[7] On average our subjects took 6.8 minutes to complete the task. Thus, the effective average hourly wage was \$13.24. We reproduce screenshots from the experiment in Appendix B.

[5] We excluded the pilot data from our subsequent analysis. However, including the data does not change the statistical significance of any of our results.

[6] See Appendix D for a breakdown of subjects' responses to these questions.

[7] Studies have found a median hourly wage of \$1.38 (Horton & Chilton, 2010) and a typical payment of \$0.01-\$0.10 per HIT (Mason & Watts, 2010).

3 Results

Table 1 and Figure 2 summarize the mean reported likelihood of punishment by treatment condition.[8] Descriptively, both our within-subject and between-subject data are in line with our hypotheses. Subjects reported a higher average likelihood of punishing A the greater was B's investment level in both the Promise and No Promise conditions, so both lines in Figure 2 are upward sloping (H.1 and H1.2). The increase in the average reported likelihood of punishing A caused by higher investment is larger when subjects were told that A made a promise to cooperate with B, so the Promise lines in Figure 2 are steeper than the No Promise lines (H1.3). Subjects reported a higher average likelihood of punishing A when subjects were told that A made a promise to cooperate with B, so the Promise lines in Figure 2 lie above the No Promise lines (H1.4).

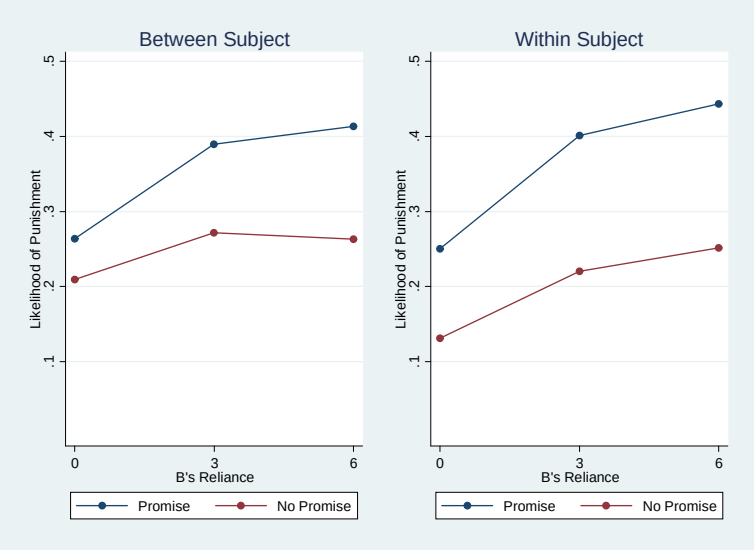


Figure 2: Average Reported Likelihood of Punishing Across Conditions

[8] We excluded the responses of the 126 subjects who reported in the post experimental survey that they did not or only "kind of" understood how A's and B's actions affected their payoffs. We also excluded the responses of four subjects who reported that they had done the study before. We have no good explanation as to why four subjects reported having taken the survey before. We provided links to participants which were only good for a single log in. We implemented filters preventing subjects (as identified by their MTurk IDs) from participating who had participated in pilots of our experiment or similar experiments we had run in the past. So the only explanation for the four self-reported repeat takers could be that subjects have multiple MTurk IDs or mistakenly checked the wrong box. Including these data does not, however, change the results. Similarly, results do not qualitatively change when we exclude subjects who report that they did not "take the scenario seriously," that they did not "carefully read the instructions," and that they chose their answers to make themselves "seem like a good person."

Table 1: Mean Reported Willingess to Punish

	Observed Investments in the Relationship		
	0	3	6
Between-Subject			
No Promise	2.47 (N=173)	2.90 (N=183)	2.84 (N=183)
Promise	2.85 (N=156)	3.73 (N=194)	3.89 (N=186)
Within-Subject (N=1075)			
No Promise	1.92	2.55	2.76
Promise	2.75	3.81	4.11

To test our hypotheses using the within-subject data we use the Wilcoxon sign-rank test, and we find support for all four hypotheses.[9] To test our hypotheses using the between-subject data we primarily use the Wilcoxon rank-sum test. In support of H1.1 we find that reliance matters if there was a promise.[10] Reliance also matters it there was no promise, but only for the (0-3) and (0-6) investment intervals.[11] This offers partial support for H1.2. For each level of investment, we find support for H1.4, our hypothesis that subjects are more willing to punish non-cooperation when there was a promise.[12] Finally, we perform a bootstrapping procedure to generate a non-parametric test of H1.3,[13] and find confirmation of our hypothesis that reliance matters more when there has been a promise.[14]

[9] We generally obtain significance at the 1% level. Only under a very restrictive definition of our data pool (eliminating all observations that might be invalid in the light of our post-experiment survey) do we find p=0.02.

[10] The effect is significant at the 1% level for the (0-3) and (0-6) intervals, and at the 5% level (p=0.02) for the (3-6) interval.

[11] We find support for H6.2 at the 1% level for the (0-3) and the (0-6) intervals. For the (3-6) interval, mean effort goes down but the effect is not significant at any level (p=0.63).

[12] The result is significant at the 1% level.

[13] The bootstrapping procedure requires us to simulate synthetic samples and is described in detail in Appendix E.

[14] The tests yields statistical significance at the 1% level for the (0-6) interval. The Wilcoxon rank-sum test is unavailable for testing H6.3 given the between-subject design. To test the interaction effect, we would have to compare the difference in the reported likelihood of punishment for different reliance levels in the Promise and the No Promise conditions. The Wilcoxon ranksum test would allow us to test for the difference between unmatched data but, as we have to test for the difference in differences, we would need within-subject data for different reliance levels to conduct this test.

4 Discussion

In Stone and Stremitzer (2017) we found evidence that the increased willingness of a promisor to keep a promise on which the promisee had relied more created an incentive for promisees to strategically overinvest in reliance in order to psychologically lock the promisor in to keeping her promise. This overinvestment disappeared when the promise was enforced with expectation damages, as the promisee could then rely on the legal regime rather than overinvestment to motivate the promisor to keep her promise. This suggests that legal enforcement not only has the potential to mitigate problems of underinvestment—the focus of the literature on breach remedies and the hold-up problem.[15] It may also reduce overinvestment arising from psychological lock-in.

In the present study we conclude that third-party observers report they are willing to punish a non-cooperative player A more in the Promise conditions than in the No Promise conditions. This willingness to punish increases with the investment B made in the relationship creating another channel for lock-in. In contrast to the mechanism posited in Stone and Stremitzer (2017), such a reputation-based lock-in mechanism doesn't rely on the internal moral motivations of the promisor. Thus, promisees should be able to harness this reputation-based lock-in mechanism by increasing their reliance investments, and so may have an incentive to overinvest even if faced with a self-interested promisor. Notice also that we find that there is something distinctively promissory about this effect, as the willingness to punish increases more with B's investments if A made a promise.

The purely hypothetical nature of the question that we posed to subjects raises a question about the external validity of our results. We might wonder whether subjects will punish in the manner that our results suggest that they will when the monetary costs of punishing are real. But the purpose of this experiment is not to provide convincing evidence that

[15] On breach remedies in particular, see, e.g., Shavell (1980, 1984), Rogerson (1984), Cooter and Eisenberg (1985), Edlin and Reichelstein (1996), Edlin (1996), Che and Chung (1999), Schweizer (2006), Ohlendorf (2009), Stremitzer (2012). On the hold-up problem generally, see, e.g., Williamson (1979, 1985), Grout (1984), Grossman and Hart (1986), Hart and Moore (1988), Chung (1991), Aghion, Dewatripont, and Rey (1994), Noeldeke and Schmidt (1995), Che and Hausch (1999).

people do in fact engage in altruistic punishment. That people exhibit such a propensity is well-supported by other work (Gintis et al., 2005). Rather, we view our results as providing evidence for the claim that to the extent that people exhibit such a propensity, they will exhibit a greater willingness to punish promise-breaking when the promisee's reliance on the promise was greater. We think a study such as ours can provide useful evidence about the relative strength of any such effect, even if the absence of monetary stakes weakens conclusions that may be drawn about the absolute magnitude of effect and thus the behavioral relevance of the propensity to altruistically punish norm violations in the first place.

5 Conclusion

Third-parties observing an interaction between a promisor and promisee are more willing to punish the promisor for breaking her promise the more the promisee has relied upon the promise. That is, third parties seem to be sensitive to the same underlying morality of promising that appears to motivate promisors: they view breaking a promise as morally worse—and so as more deserving of punishment—the more the promisee has relied upon the promise. This suggests that reputational forces may provide a promisee with a mechanism for locking in the promisor who isn't herself motivated by moral forces: by investing more, the promisee increases the chance that third-parties who are sensitive to these moral forces will punish the promisor should she break the promise.

Our results also highlight a more general point. Studying the reasons why people voluntarily keep their promises sheds light on the dynamics of altruistic punishment, and, we conjecture, the dynamics second-party enforcement mechanisms that are embedded in relational contracts. The norms that make a promisor keep a promise also make seem to inform the behavior of those who can punish promise breakers. This suggests that there is a stronger link between relational contracting, altruistic punishment, and voluntary promise keeping than has been noticed in the contract theory literature to date.

References

Aghion, P., M. Dewatripont, and P. Rey (1994): "Renegotiation Design with Unverifiable Information," *Econometrica*, 62(2), 257–282.

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

——— (2010): "Bare Promises: An Experiment.," *Economics Letters*, 107, 281–83.

Che, Y. K., and T. Y. Chung (1999): "Contract Damages and Cooperative Investments," *RAND Journal of Economics*, 30(1), 84–105.

Che, Y. K., and D. B. Hausch (1999): "Cooperative Investments and the Value of Contracting," *American Economic Review*, 89(1), 125–147.

Chung, T. Y. (1991): "Incomplete Contracts, Specific Investments, and Risk Sharing," *Review of Economic Studies*, 58(5), 1031–1042.

Dufwenberg, M., and U. Gneezy (2000): "Measuring Beliefs in and Experimental Lost Wallet Game," *Games and Economics Behavior*, 30(2), 163–182.

Ederer, F., and A. Stremitzer (2017): "Promises and expectations," *Games and Economic Behavior*, 106, 161–178.

Edlin, A. S. (1996): "Cadillac Contracts and Up-front Payments: Efficient Investment Under Expectation Damages," *Journal of Law, Economics, and Organization*, 12(1), 98–119.

Edlin, A. S., and S. Reichelstein (1996): "Holdups, Standard Breach Remedies, and Optimal Investment," *American Economic Review*, 86(3), 478–501.

Ellingsen, T., and M. Johannesson (2004): "Promises, Threats and Fairness," *Economic Journal*, 114(495), 397–420.

Fehr, E., and U. Fischbacher (2004): "Third-party punishment and social norms," *Evolution and human behavior*, 25(2), 63–87.

Gintis, H., S. Bowles, R. Boyd, and E. Fehr (2005): "1 Moral Sentiments and Material Interests: Origins, Evidence, and Consequences," *Moral sentiments and material interests*, p. 1.

Grossman, S. J., and O. D. Hart (1986): "The costs and benefits of ownership - A theory of vertical and lateral integration," *Journal of Political Economy*, 94(4), 691–719.

Grout, P. (1984): "Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach," *Econometrica*, 52 (2), 449–460.

Hart, O., and J. Moore (1988): "Incomplete Contracts and Renegotiation," *Econometrica*, 56(4), 755–785.

Hart, O., and J. Moore (2008): "Contracts as reference points," *The Quarterly journal of economics*, 123(1), 1–48.

Horton, J. J., and L. B. Chilton (2010): "The Labor Economics of Paid Crowdsourcing," in *Proceedings of the 11th ACM Conference on Electronic Commerce*, EC '10, pp. 209–218, New York, NY, USA. ACM.

Mason, W., and D. J. Watts (2010): "Financial incentives and the performance of crowds," *ACM SigKDD Explorations Newsletter*, 11(2), 100–108.

Mischkowski, D., R. Stone, and A. Stremitzer (2016): "Promises, Expectations, and Social Cooperation," *Harvard Law School John M. Olin Center Discussion Paper No. 887*.

Nöldeke, G., and K. M. Schmidt (1995): "Option Contracts and Renegotiation - a Solution to the Hold-Up Problem," *RAND Journal of Economics*, 26(2), 163–179.

Ohlendorf, S. (2009): "Expectation Damages, Divisible Contracts, and Bilateral Investment," *American Economic Review*, 99 (4), 1608–1618.

Rogerson, W. P. (1984): "Efficient Reliance and Damage Measures for Breach of Contract," *RAND Journal of Economics*, 15(1), 37–53.

Schweizer, U. (2006): "Cooperative Investments Induced by Contract Law," *RAND Journal of Economics*, Spring 37-1, 134–145.

Shavell, S. (1980): "Damage Measures for Breach of Contract," *Bell Journal of Economics*, 11.

Stone, R., and A. Stremitzer (2017): "Promises, Reliance, and Psychological Lock-in," *UCLA School of Law, Law-Econ Research Paper No. 15-17*.

Stremitzer, A. (2012): "Standard Breach Remedies, Quality Thresholds, and Cooperative Investments," *Journal of Law, Economics, and Organization*, 28(2), 337–359.

Vanberg, C. (2008): "Why Do People Keep Their Promises? An Experimental Test of Two Explanations.," *Econometrica*, 76, 467–1480.

Williamson, O. E. (1979): "Transaction-Cost Economics: The Governance of Contractual Relations," *Journal of Law and Economics*, 22, 233–61.

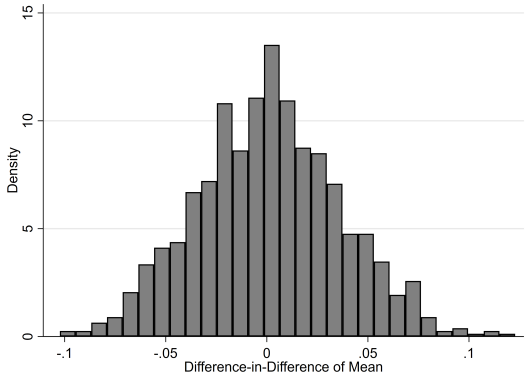
——— (1985): *The Economic Institutions of Capitalism*. Free Press, New York.

APPENDIX A: BOOTSTRAP

We performed a two-step bootstrapping procedure in order to generate a non-parametric test of H1.4 on the (0-6) gap. Let $\hat{\theta}_r$ and θ_r be the estimators of the mean reported likelihood of punishment in the Promise and No Promise conditions respectively when the promisee's reliance investments are r. We observe that $\left(\hat{\theta}_6 - \theta_6\right) - \left(\hat{\theta}_0 - \theta_0\right) = 0.67$. That is, the difference in means between the Promise and the No Promise samples is higher if promisees' reliance investments are 6 than if they are 0. We want to know the probability with which we would observe this positive difference-in-difference of means by chance. In other words, we want to test the null hypothesis that $\left(\hat{\theta}'_6 - \theta'_6\right) - \left(\hat{\theta}'_0 - \theta'_0\right) = 0$, where $\hat{\theta}'_1, \theta'_1, \hat{\theta}'_0, \theta'_0$ are the means of the underlying distributions from which our samples are drawn.

We tested the null hypothesis in two steps (see, e.g., Efron and Tibshrani, 1993, pp. 220-223). First, we recentered the original samples to conform with the null hypothesis. Specifically, we subtracted from each observation in each of the four samples the respective sample means and then added to each observation in the two Promise samples the mean effect of promising. In other words, if the mean for the combined No Promise samples is $\bar{x}$ and the mean for the combined Promise samples is $\bar{y}$, we added $(\bar{y} - \bar{x})$ to each observation in the two Promise samples.[16]

We then created four synthetic samples with sample sizes equal to our real samples by randomly drawing with replacement from each of the four samples that were constructed above. We then calculated the difference-in-difference $\left(\hat{\theta}''_6 - \theta''_6\right) - \left(\hat{\theta}''_0 - \theta''_0\right)$, where $\hat{\theta}''_1, \theta''_1, \hat{\theta}''_0, \theta''_0$ are the means of these synthetic samples. After repeating this procedure 10,000 times, we obtained a simulated distribution of the differences-in-differences of the means that would arise if the null hypothesis were true (that is, if the difference of means between the Promise and the No Promise conditions was equal across different reliance levels).



Simulated Distribution for the (0-6) gap

[16] By subtracting the sample means, we made our data conform to the hypothesis $\hat{\theta}'_6 = \theta'_6 = \hat{\theta}'_0 = \theta'_0 = 0$. In doing so, we eliminated all of our hypothesized effects from our data. By adding back $(\bar{y} - \bar{x})$ to the observations in the promise samples, we effectively added back in the expectations-independent effect of promising, so that our data ended up conforming to our less restrictive null hypothesis $\left(\hat{\theta}'_6 - \theta'_6\right) - \left(\hat{\theta}'_0 - \theta'_0\right) = 0$. We didn't add back in the effect of reliance alone, because doing so would leave this hypothesis unchanged.

The area under the curve to the right of the observed estimator 0.67 corresponds to the probability that a greater or equal difference-in-difference would have been observed if the null hypothesis were true. This value, 0.004, is small enough for us to reject the null hypothesis at the 1% level.

APPENDIX B: INSTRUCTIONS