PROMISES, RELIANCE, AND PSYCHOLOGICAL LOCK-IN

BY

REBECCA STONE
UCLA SCHOOL OF LAW

ALEXANDER STREMITZER
ETH ZURICH

In 49 JOURNAL OF LEGAL STUDIES 33 (2020)

Promises, Reliance, and Psychological Lock-In*

Rebecca Stone† UCLA

Alexander Stremitzer‡ ETH Zurich

May 1, 2020

Abstract

We obtain experimental evidence that suggests: (i) promisors are more likely to keep promises the more they were relied upon by a promisee; (ii) this effect is anticipated by promisees, who accordingly strategically overinvest to lock promisors into keeping their promises ("psychological lock-in"); and therefore (iii) legal enforcement can reduce overinvestment as promisees don't need to make use of the extra-legal mechanism of psychological lock-in when the legal regime induces promisors to perform instead. These results contrast with the central prediction of the holdup literature that underinvestment results in the absence of legal enforcement.

Keywords: promises, reliance, remedies, contracts, crowding out.

JEL-Classification: K12, A13, C91, C72, D64.

1 Introduction

In the absence of legal enforcement, a promisee (a recipient of a promise) may be wary of relying on a promise for fear that the promisor (the maker of the promise) won't keep it. Because classical economic theory assumes that agents are rational and self-interested,

*For helpful comments we are grateful to the editor, William Hubbard, an anonymous referee, Yonathan Arbel, Oren Bar-Gill, Cristina Bicchieri, Patrick Bolton, Gary Charness, Bob Cooter, Joseph Doherty, Florian Ederer, Gillian Hadfield, Rob MacCoun, Andreas Roider, Bob Scott, Seana Shiffrin, Kathy Spier, and Eric Talley. We are also grateful to seminar audiences at Alabama, Berkeley, Boston University, Columbia, ETH Zurich, the University of Zurich, Harvard, Paris II, UCLA, University of Pennsylvania, USC, the 2015 EMLS Workshop held at UCLA, the ISNIE 2015 meeting, the Annual Conference of the Polish Law and Economics Society, and the Latin American Law & Economics Meeting. We thank Calvin Cam, Shane Farley, Atif Ghulam, Rita Hsu, Henry Kim, Brandon Milostan, and Reggie Young for excellent research assistance. Financial Support from a 2014-15 UCLA Faculty Research Grant is gratefully acknowledged.

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

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

it predicts that a promisee will underinvest in reliance on a promise whenever he believes it likely that the promisor will have a self-interested reason to break it. Legal enforcement of promises should accordingly mitigate this problem of underinvestment by giving the promisor a self-interested reason to keep her promise, thus assuring the promisee that his investment won't be wasted. A large literature on breach remedies (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) and the holdup problem (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) studies how the introduction of third-party enforcement can mitigate this underinvestment problem.[1]

Yet reducing underinvestment might not be the only virtue of legal enforcement if some people are intrinsically motivated to keep their promises. Casual observation, introspection, and plenty of empirical evidence tell us that many people are motivated to keep their promises, even in the absence of self-interested reasons to do so (Ellingsen and Johannesson 2004; Charness and Dufwenberg 2006; Vanberg 2008; Charness and Dufwenberg 2010; Ostrom, Walker, and Gardner 1992; Kerr and Kaufman-Gilliland 1994; Sally 1995; Bicchieri and Lev-On 2007). There is also evidence that promisors are more likely to keep their promises the more they believe that promisees expect them to do so (Mischkowski, Stone, and Stremitzer 2019; Ederer and Stremitzer 2017). Given the close connection between expectations of performance and actions taken by the promisee in reliance on a promise, a promisor's intrinsic motivation to keep a promise is likely to also increase the more the promisee has relied upon it.[2] If this is the case, then the promisee has an incentive to strategically rely

[1] 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).

[2] Consistent with this conjecture, Wilkinson-Ryan and Hoffman (2015) provide evidence based on a psychological vignette study that people are more likely to keep relied upon promises. However, in their study such reliance confers a material benefit on the promisor. Thus, they don't disentangle the effects of reliance alone from the effects of reciprocity.

on the promise in order to make the promisor more likely to keep it—an effect that we refer to as "psychological lock-in." Indeed, in the absence of a legal regime, the promisee might *overinvest* in order to psychologically lock-in the promisor. Legal enforcement of relied-upon promises would then reduce a promisee's need to overinvest in reliance on a promise, because the legal regime would motivate the promisor to keep her promise instead. Legal enforcement might therefore have the unexpected benefit of reducing overinvestment alongside the more expected benefits of reducing underinvestment and increasing rates of promise keeping.

In this paper, we experimentally investigate three questions in the laboratory. First, does a promisee's greater reliance on a promise make the promisor more likely to keep it? Second, does the promisee anticipate such an effect and so strategically increase his reliance to make the promisor more likely to keep her promise? Third, what benefits for social welfare arise from the legal enforcement of a promise? In particular, in what ways does such legal enforcement improve the promisee's investment decision? Could it be that legal enforcement reduces overinvestment by reducing the promisee's need to avail himself of the extra-legal mechanism of psychological lock-in to motivate the promisor to keep her promise?

The form of legal regime that we study in this paper mimics a legal regime that is governed by the doctrine of promissory estoppel. In its canonical form, the doctrine allows a promisee to legally enforce a gratuitous promise if the promisee detrimentally relied on the promise and this reliance was reasonably foreseeable to the promisor (Restatement (2d) of Contracts § 90, 1981). Gratuitous promises are made without the promisor getting something in return and therefore don't satisfy the consideration requirement, the usual test for the enforceability of promises in American contract law.

The earliest applications of promissory estoppel were confined to promises that were not motivated by the prospect of commercial gain.[3] However, the domain of the doctrine has subsequently grown considerably (Holmes 1996). Courts now allow reliance to act as

[3]The historical roots of the doctrine of promissory estoppel can be traced to a few categories of gratuitous promises: (1) gratuitous promises to convey land; (2) gratuitous promises by bailees; (3) promises to make gifts to charitable institutions; and (4) gratuitous promises among family members. (Farnsworth 2004, pp. 90-92)

a consideration substitute in many situations in which the promisor is motivated by the prospect of commercial gain. For example, gratuitous firm offers, voluntary waivers, and one-sided modifications, which once could be freely repudiated for lack of consideration, are now binding obligations when relied upon by the promisee (Knapp 1998, p. 1198).[4]

Moreover, contemporary courts sometimes invoke promissory estoppel to enforce promises which are supported by consideration, when those promises would be rendered unenforceable for some other reason, for example, a lack of definiteness or a failure to satisfy the Statute of Frauds. Thus, in *Wheeler v. White* (398 S.W.2d 93 [1966]), White promised to obtain a loan for Wheeler to enable Wheeler to construct a commercial building on a tract of land owned by Wheeler. In return, Wheeler promised future payments including commissions on rentals received from any tenants of the new building. Following reassurances by White that the loan would be forthcoming, Wheeler proceeded to raze the existing building and prepare the land for the new structure. White then informed Wheeler that no loan would be forthcoming. In the subsequent litigation, Wheeler's contract claim was denied on the grounds that the agreement was too indefinite to be enforceable. But the Supreme Court of Texas ultimately enforced White's promise on promissory estoppel grounds.[5] In a similar vein, in *Alaska Airlines, Inc. v. Stephenson* (217 F.2d 295 [1954]), Stephenson gave up his job with Western Airlines after securing an oral agreement on a two-year employment contract with Alaska Airlines. Alaska Airlines then fired Stephenson before the two years were up and argued that the agreement was unenforceable because it hadn't been formalized in a writing signed by Alaska Airlines and therefore failed to satisfy the requirements of the Statute of Frauds. The court, however, found Alaska's promise of employment enforceable because of Stephenson's foreseeable reliance on it.[6]

[4]Courts have also allowed promissory estoppel claims based on promises of at-will employment to proceed where the prospective employee has relied on such a promise by, for example, moving her family a long distance or giving up other sources of income or employment. See, e.g., *Grouse v. Group Health Plan, Inc.* (306 N.W.2d 114 [1981]); *Ravelo v. County of Hawaii* (658 P.2d 883 [1983]).

[5]For other examples of cases that invoke promissory estoppel to enforce contracts that would be unenforceable on grounds of definiteness, see *Rosnick v. Dinsmore* (457 N.W.2d 793 [1990]); *Hoffman v. Red Owl Stores, Inc.* (133 N.W.2d 267 [1965]).

[6]For other examples of cases that invoke promissory estoppel to enforce contracts that would be unenforceable pursuant to the Statute of Frauds, see *Kolkman v. Roth* (656 N.W.2d 148, 153 [2003]); *Olympic*

Our experiment mimics the structure of cases like *Alaska Airlines* and *Wheeler*. At the outset of our experiment, pairs of subjects were given the opportunity to exchange promises to cooperate in a subsequent experimental stage. That is, they could exchange promises to do something valuable for each other—as Alaska Airlines and Stephenson did in entering into their employment agreement, and White and Wheeler did in forming their agreement concerning the financing of Wheeler's building project. Some of our subjects were then assigned to a treatment in which they faced no sanctions for breaking their promise, while others were assigned to a treatment that models a promissory estoppel regime: subjects who broke their promises had to compensate the promisee, but only if the promisee had relied upon the promise—just as Stephenson's resignation from his job with Western Airlines and Wheeler's preparation of this land for the new building served as the grounds for enforcing Alaska Airline's and White's promises.

Although canonical promissory estoppel cases involve purely gratuitous promises, we designed our experiment to model exchanges of promises because cases like *Alaska Airlines* and *Wheeler* are likely to have greater commercial importance and remain controversial as a doctrinal matter thus increasing their significance from a policy standpoint.[7] Thus, our experimental results don't directly speak to applications of the doctrine to purely gratuitous promises. However, our results should generalize to such settings so long as people don't distinguish among promises according to the motive with which they are made.

Judges have discretion to determine the remedy once they have found a promisor liable on promissory estoppel grounds. In particular, they can choose between expectation damages—the standard remedy for breach of contract, which is designed to make the promisee as well off as he would have been had the promise been kept—and reliance damages—a remedy designed

Holding Co. v. ACE Ltd. (909 N.E.2d 93, 94 [2009]); *Jamestown Terminal Elevator, Inc. v. Hieb* (246 N.W.2d 736 [1976]). See also Restatement Second of Contracts § 139(1) (1981).

[7]On the controversy surrounding the application of promissory estoppel to indefinite agreements, see, e.g. *Keil v. Glacier Park, Inc.* (614 P.2d 502, 506 [1980]); *Black Canyon Racquetball Club, Inc. v. Idaho First Nat. Bank* (804 P.2d 900, 907 [1991]); *Forstmann v. Culp* (648 F.Supp. 1379 [1986]); see also Scott (2007). For expressions of skepticism about the application of promissory estoppel in the statute of frauds context, see *Tanenbaum v. Biscayne Osteopathic Hospital, Inc.* (190 So.2d 777, 779 [1966]); *Mullins v. Southern Pacific Transp. Co.* (851 P.2d 839 [1992]); *Genin, Trudeau & Co. v. Integra Development International* (845 F.Supp. 622 [1994]).

to ensure that the promisee is no worse off than he would have been had he never received and relied on the promise (See Restatement (2d) of Contracts § 90, 1981). But the scholarly consensus seems to be that judges usually award expectation damages in promissory estoppel cases (Farber and Matheson 1985; Yorio and Thel 1991). Thus, in this paper, we study the effects of introducing a promissory estoppel regime with expectation damages (Expectation Damages). We do so by contrasting it with a regime in which legal enforcement is absent (No Regime).[8]

Consistent with the predictions of a simple model, we find that in the absence of legal enforcement, promisors are more likely to cooperate with promisees the more promisees relied on their promises and that promisees believe promisors to act in this way. And we find that the introduction of a promissory estoppel regime with expectation damages decreases overinvestment. These findings constitute evidence of a psychological lock-in effect.

The remainder of the paper is organized as follows. Section 2 reviews related literature. Section 3 describes the experimental design and procedures, outlines the model of agents' preferences that forms the basis of our predictions, and formulates the specific hypotheses that we test. Section 4 presents our results, which are largely in line with our theoretical predictions. We discuss potential limitations of our results in Section 5 and implications for legal theory in Section 6. Section 7 concludes.

2 Related Literature

Two theories explaining why many promisors are intrinsically motivated to keep their promises have dominated the recent experimental literature. Proponents of the *expectation-based* theory argue that promisors keep their promises in order to avoid failing to meet the expectations of performance that their promises have created in the promisee (Dufwenberg and Gneezy 2000; Charness and Dufwenberg 2006). In contrast, the *commitment-based* theory claims that promisors have a preference for keeping their word regardless of the promisee's expec-

[8]In *Alaska Airlines v. Stephenson* the court affirmed the lower court's award of expectation damages, while the court in Wheeler v. White held that Wheeler was entitled to reliance damages.

tations (Ostrom, Walker, and Gardner 1992; Braver 1995;; Ellingsen and Johannesson 2004; Vanberg 2008; Ismayilov and Potters 2012).

It is, of course, possible that *both* a promisee's expectations of performance and a promisor's sense that she is duty-bound to keep her promise regardless contribute to her willingness to keep her promise. Vanberg (2008), argues that reasons to honor their commitments regardless are the *only* reasons why people keep their promises. Similarly, Ellingsen et al. (2010) find no significant relationship between expectations and contributions in dictator and trust games in which they elicited expectations from recipients and communicated those expectations to the dictators and trustees respectively. However, in both of theses studies, the findings that expectations play no role arise under conditions in which the dictator (or trustee) had not made a promise to the recipient (or trustor).[9]

In Charness's and Dufwenberg's (2006) experiment, by contrast, there were such promises, and the results suggest that expectations matter. But these results are based on correlations between second-order beliefs and actions that allow for alternative explanations. Ederer and Stremitzer (2017) get around this problem by exploiting exogenous variation in promisors' second-order beliefs. They show that promisee expectations affect promise keeping in cases in which expectations are supported by direct promises. And they postulate that motivations to keep promises may depend on an interaction effect: a promisee's expectations matter if and only if the promisor actually made a promise to the promisee. Using a vignette design, Mischkowski, Stone, and Stremitzer (2019) find evidence of a weaker interaction effect, showing that promising makes promisors more sensitive to promisees' expectations.

We assume that a promisor's motivation to keep her promise increases as the promisee's reliance upon the promise increases. This might be because a promisee's reliance is an indicator of his expectations that the promisor will perform, and the promisor dislikes disappointing the promisee's expectations. But it also might be because the promisor simply

[9]Vanberg finds that exogenous variation of promisors' second-order expectations—that is, their beliefs about promisees' beliefs—doesn't generate a statistically significant difference in promise keeping. But in his design, subjects were rematched after a round of promises was made, so that promisors ultimately faced promisees to whom they made no promises. In Ellingsen et al.'s study, no promises were ever made.

doesn't like to make the promisee worse off than he would have been had the promise been kept. This implies that the more the promisee relied upon the promise, the more willing the promisor will be to keep it.[10]

Our paper is also related to the experimental literature that examines the effects of legal enforcement on investment incentives. Sloof et al. (2003) and Sloof et al. (2006) obtain experimental evidence that, consistent with the predictions of classical economic theory (e.g., Shavell 1980), shows that enforcement of contracts with expectation damages protects a promisee's investment too well by encouraging overinvestment, both when renegotiation is not possible (Sloof et al. 2003) and when renegotiation is possible (Sloof et al. 2006). In their design, there is a chance that performance of the contract might be inefficient. Overinvestment occurs because expectation damages perfectly insure the promisee against the risk of breach, even in states of the world in which performance of the contract is inefficient. The promisee disregards this possibility when choosing his investment level and so invests the amount that would be desirable if performance were always efficient.[11] Our design abstracts from this reason to overinvest by ensuring that promise keeping is always efficient.[12]

3 Design and Procedure

In this section, we describe our experimental design and procedures, and the hypotheses that arise from our simple model.

[10]We focus on material harms promisees suffer as a result of their reliance on promises, as opposed to the psychic harms that may be associated with disappointed expectations. Thus, for simplicity, we don't assume that a promisor's second-order beliefs about the promisee's expectations influence her preferences. This means that, unlike the expectation theorists, we don't need to employ the apparatus of Battigalli's and Dufwenberg's (2007, 2009) psychological game theory model.

[11]Sloof and his coathors also show that, consistent with theoretical predictions, reliance damages cause even more overinvestment. Reliance damages, like expectation damages, perfectly insure the investment decision against the possibility of breach. But reliance damages give the investor an additional incentive to invest to reduce the likelihood of breach.

[12]This is not an unrealistic assumption in a relevant subset of cases, as we explain in Section 6.

3.1 Experimental Design

Subjects played a modified-dictator game in which a Dictator decides whether to cooperate with a Recipient by sending money to him. The game differs from a standard dictator game in two ways. First, at the outset of the game players don't know whether they will be assigned the role of the Dictator or the Recipient. Behind this veil of ignorance they decide whether to exchange promises to cooperate if assigned the role of Dictator during a "communication stage." Second, once roles have been assigned, but before the Dictator decides whether to cooperate ($a = 1$) or not ($a = 0$), the Recipient chooses a level of investment, $i \in \{0, 1, 2, 3, 4, 5, 6\}$.[13]

In our No Regime scenario, which is depicted in Figure 1, the Recipient's monetary payoff first increases and then decreases in his investment if the Dictator cooperates, while it always decreases his payoff and does so at a faster rate if the Dictator doesn't cooperate. Regardless of the Dictator's action, the Recipient's investment has no effect on the monetary payoff of the Dictator. Thus, the Dictator has no reciprocity-based reason to reward the Recipient for a positive investment choice. Specifically, the Dictator's monetary payoff is given by

$$\pi_D(a, i) = \begin{cases} 15 & \text{if } a = 0 \\ 12 & \text{if } a = 1 \end{cases}, \tag{1}$$

and the Recipient's by

$$\pi_R(a, i) = \begin{cases} 6 - i & \text{if } a = 0 \\ 12.25 - 0.25\,|1 - i| & \text{if } a = 1 \end{cases}. \tag{2}$$

If the Dictator cooperates, the Recipient maximizes his monetary payoff by investing one. If the Dictator doesn't cooperate, the Recipient maximizes his payoff by investing zero. As in a standard dictator game, the Dictator maximizes her monetary payoff by choosing not to cooperate.

[13] Here we describe the version of the game we implement in our experiment. For full analysis of the more general model on which the game is based, see Online Appendix A.

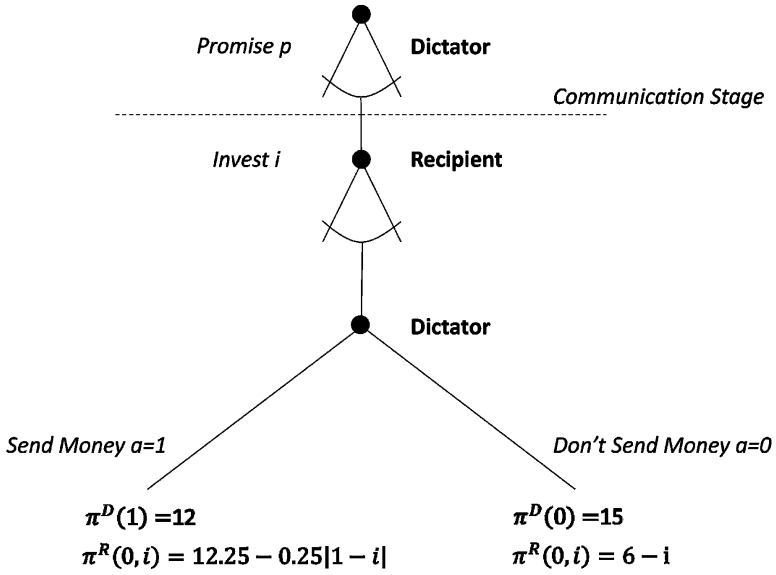


Figure 1: Game Tree in No Regime.

We compare No Regime with a scenario, Expectation Damages, in which promise keeping is enforced with expectation damages so long as the Recipient has invested something in reliance on any promise to cooperate that the Dictator made him during the communication stage. Monetary payoffs are unchanged if the Dictator keeps her promise or made no promise, but if she breaks a promise and the Recipient invested at least one, then the Dictator must pay the Recipient an amount of money $l(i)$ that ensures that his monetary payment is the same as if the Dictator had kept her promise. In this way, as we explained in the introduction, we model applications of the doctrine of promissory estoppel to agreements that are supported by consideration, but unenforceable for some other reason like lack of definiteness or a failure to satisfy the Statute of Frauds. Thus,

$$l(i) = \pi_R^L(0,i) - \pi_R(0,i) = \begin{cases} 0 & \text{if } i = 0 \\ 6.5 - 0.75i & \text{if } i > 0 \end{cases}, \tag{3}$$

and when the Dictator breaks a promise to cooperate, the Recipient's payoffs are given by

$$\pi_R^L(0,i) = \pi_R(0,i) + l(i) = \begin{cases} 6 & \text{if } i = 0 \\ 12.50 - 0.25i & \text{if } i > 0 \end{cases}, \tag{4}$$

and the Dictator's by

$$\pi_D^L(0,i) = 15 - l(i) = \begin{cases} 15 & \text{if } i = 0 \\ 8.50 - 0.75i & \text{if } i > 0 \end{cases}. \tag{5}$$

Predictions of the standard model. The standard self-interested rational actor model assumes that each subject maximizes his monetary payoff. It therefore predicts that the Dictator will never cooperate in No Regime (because $\pi_D(0) = 15 > 12 = \pi_D(1)$) and that, anticipating this choice by the Dictator, the Recipient will invest zero. By contrast, in Expectation Damages, the Dictator will always cooperate if and only if the Recipient invested at least one (because $\pi_D(1) = 12 > 8.5 - 0.75i = \pi_D^L(0, i)$). Thus, the Recipient will invest one as this maximizes his payoff in the event that the Dictator cooperates, thus guaranteeing the socially optimal outcome.

Intrinsic motivation to keep promises. Consistent with prior evidence, we assume, contrary to the standard model, that subjects are intrinsically motivated to keep their promises. Thus, we suppose that the Dictator cares both about her material payoff and keeping any promise she made to the Recipient and that she cares more about keeping a promise the more the Recipient would lose should she break it. We therefore posit that, if the Dictator makes a promise, $p = 1$, and subsequently breaks it, $a = 0$, her utility is reduced by an amount that depends on the resulting reduction of the Recipient's material payoff.[14] In line with the economics literature on promising, we refer to the latter component of the utility as the Dictator's "guilt" (Charness and Dufwenberg 2006; Battigalli and Dufwenberg 2007, 2009; Ederer and Stremitzer 2017).[15] Formally, we suppose that the Dictator's utility is given by:

$$u_D(a, p) = \pi_D(a) - p\gamma g\left(\pi_R(1, i) - \pi_R(a, i)\right), \tag{6}$$

[14]Because our focus is on the effects of reliance on promising, we abstract from other considerations that may drive agents to keep their promises like a desire not to disappoint promisees' expectations or a simple desire to do as one promised (Mischkowski, Stone, and Stremitzer 2019).

[15]Although we follow the literature in using the terminology of "guilt,"we are agnostic over whether the "guilt" function should be given a literal interpretation. The use of the term "guilt" might be taken to suggest that the Dictator's psychological well-being is reduced when she breaks a promise, thus giving the Dictator a self-interested reason to keep a promise. In this case, "guilt" should enter the social welfare function as it would represent a component of the Dicator's welfare. But we could instead regard it simply as capturing the strength of the non-self-interested considerations that the Dictator believes give her reasons to keep her promises. If so, then "guilt" would not represent a component of the Dictator's welfare, and so ought not to be included in the social welfare function.

where $\gamma \geq 0$ is a parameter that represents the weight she placed on her guilt relative to her material payoff, and $g(x)$ is the guilt function with $g(0) = 0$ and $g'(x) > 0$. Guilt is therefore zero whenever the Dictator made no promise, $p = 0$, or whenever she cooperates, $a = 1$: $u_D(a, 0) = \pi_D(a)$ and $u_D(1, 1) = \pi_D(1)$. Because the Recipient made no promise to choose a particular investment level, such a motivation has no relevance to his decision making, and so we assume that his utility function is simply his payoff function: $u_R(a, i) = \pi_R(a, i)$.

Our design is a particular instantiation of a more general model that we solve fully by backwards induction in Online Appendix A. Here we highlight the features of the equilibrium that form the basis of our hypotheses starting with No Regime. The central observation is that in the absence of a material penalty for breaking a promise, the guilt function may provide the Dictator with sufficient inducement to keep her promises. A Dictator who made no promise to cooperate will never find it in her interests to cooperate because $\pi_D(0) = 15 > 12 = \pi_D(1)$. But a Dictator who made a promise will cooperate whenever

$$\pi_D(1) \geq \pi_D(0) - \gamma g(\pi_R(1, i) - \pi_R(0, i)), \tag{7}$$

that is, whenever the sensitivity to guilt, γ, exceeds a critical value, $\gamma^c(i)$, which depends on the Recipient's investment level:

$$\Leftrightarrow \gamma \geq \gamma^c(i) = \frac{\pi_D(0) - \pi_D(1)}{g(\pi_R(1, i) - \pi_R(0, i))}. \tag{8}$$

Because the guilt function, $\pi_R(1, i) - \pi_R(0, i)$, is increasing in i, the guilt the Dictator experiences when she breaks a promise increases in investment, and the critical value $\gamma^c(i)$ that determines whether she will cooperate decreases.

It follows that when γ is so low that even the highest possible investment level won't trigger enough guilt to induce the Dictator to cooperate, that is, $\gamma < \gamma^c(6)$, then the Dictator will never cooperate and a Recipient, if he knows γ, will accordingly invest zero. Conversely, when γ is so high that investing the efficient amount is sufficient to induce the Dictator to cooperate, $\gamma > \gamma^c(1)$, then the Recipient will chose the efficient investment level, and the Dictator will cooperate, and social welfare is maximized. Finally, when γ is in an

intermediate range, whether the Dictator will be motivated to keep her promises depends on the Recipient's investment level, but investing the efficient amount is not sufficient to induce the Dictator to cooperate. Thus, the Recipient will choose a level of investment greater than one in order to make the Dictator's guilt from breaking the promise sufficient to induce the Dictator to keep it—that is, to psychologically lock-in the Dictator.[16] Accordingly, guilt can also produce a new kind of inefficiency: overinvestment. The amount the Recipient needs to overinvest in this intermediate range is decreasing in γ, as the higher is the Dictator's internal motivation to keep her promise, the less the Recipient needs to increase the guilt function, $\pi_R(1,i) - \pi_R(0,i)$, in order to motivate the Dictator to keep it. Figure 2 illustrates how equilibrium investment changes assuming a linear guilt function and a Recipient who knows γ.

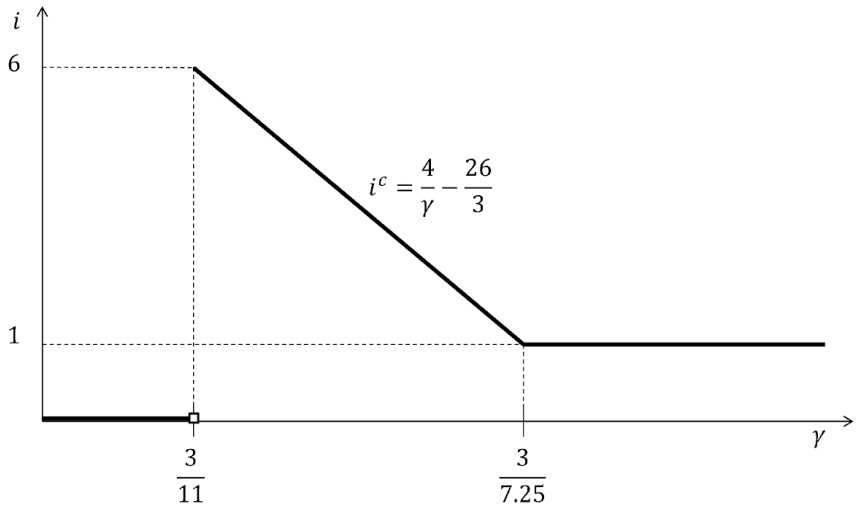


Figure 2: Equilibrium investment as function of γ

In Expectation Damages, the enforcement regime ensures that the Recipient is always as well off when the Dictator cooperates as when he doesn't, so long as the Dictator made a promise to cooperate and the Recipient invests at least one. Thus, although the Dictator won't experience guilt from breaking a promise in this case, by investing at least one the Recipient can ensure that any Dictator, regardless of her guilt parameter, will prefer to

[16]This result obtains because when the Recipient invests the amount required to induce the Dictator to cooperate, his payoff is greater than the payoff he would receive were he to invest zero. If this condition did not obtain, then the Recipient would choose a zero investment level and the Dictator would not cooperate.

cooperate. And so, in equilibrium the Recipient will invest one so long as the Dictator made a promise, leading to the socially optimal result. Hence, the equilibrium prediction of our model is the same as that of the standard model.

Working backwards, we can now analyze the communication stage. Under Expectation Damages, both parties always exchange promises in equilibrium. This is because, behind the veil of ignorance about the parties' roles in the subsequent subgame, the benefit of receiving a promise in return outweighs the costs of making one. The same applies in No Regime so long as each party believes that the other has a guilt parameter large enough to induce her to keep her promise in the subsequent subgame. The communication stage is analyzed in more detail in Online Appendix A.

We also compare No Regime to a third scenario, Control, which is identical to No Regime, except there is no communication stage. Because promising is therefore not possible in Control, the comparison enables us to assess the extent to which subjects' promises are driving their behavior in No Regime.

3.2 Experimental Procedure

Subjects played several rounds of one of the games described above. At the outset of each round, each subject was randomly and anonymously matched with another subject. The round then consisted of four steps.

Step 1. During the communication stage, subjects exchanged pre-coded messages with one another sequentially in a randomly determined order. The subject who was selected to make the first communication decision (Participant 1) had to decide whether to send Message 1 to the subject with whom he had been paired (Participant 2): "I promise to send you money if I am chosen to be Player A so long as you make me a return promise." If Participant 1 decided not to send Message 1 to Participant 2, she instead sent him a message that said "I don't promise." After learning of Participant 1's decision, Participant 2 then had to decide whether to send Message 2 to Participant 1: "I promise to send you money if I am chosen to be Player A." If Participant 2 decided not send Message 2 to Participant 1,

he instead sent her a message that said "I don't promise." Notice that the conditional form of Message 1 means that when Participant 1 sent Message 1, she only ended up promising to send Participant 2 money if Participant 2 made her a promise in return.[17] Figure 8 in the Appendix shows which subjects ended up making a promise depending on the messages they sent one another during the communication stage. Given this message structure, our model predicts that subjects only make non-gratuitous promises in equilibrium. That is, they only make promises to cooperate when they anticipate that doing so will elicit a promise in return.[18]

Step 2. Once this communication stage was over, subjects' roles for the round were selected, and subjects learned whether they would be the Dictator (Player A) or the Recipient (Player B). Roles were randomly assigned anew in each round.

Step 3. The Recipient then indicated his level of confidence that the Dictator would cooperate with him for each possible investment level and chose how much to invest by selecting a number between zero and six.

Step 4. Once the Recipient had made his investment decision, the Dictator had to decide whether to cooperate with the Recipient by sending him money. We used the strategy method to elicit the Dictator's cooperation decision (Selten 1967). That is, the Dictator made her decision in ignorance of the Recipient's actual choice by indicating whether she wanted to cooperate for every possible investment level that the Recipient might have chosen. The Dictator's actual choice, and therefore the final payoffs of both players, was given by the Dictator's choice for the Recipient's chosen investment level. If, for example, the Recipient

[17] Vanberg (2008) and Hoppe and Schmitz (2011) also use endogenously generated promises, while Sloof et al. (2003) and Sloof et al. (2006) simply tell their subjects to imagine that they had signed a contract. Vanberg (2008) uses free-form messages, allowing subjects to exchange a sequence of instant messages, instead of computer-coded messages. Free-form messages may be more natural, possibly explaining why they produce higher effect sizes (see, e.g., Charness and Dufwenberg 2010). But they have the disadvantage that subjects may make promises without using promissory language, making it difficult to disentangle promises from mere statements of intent.

[18] See Section 3.1 and Online Appendix A. Thus, our design gives subjects good (self-interested) reasons to make promises, and most end up making promises, which is important given that our aim is to study determinants of promise keeping. Furthermore, had subjects been free to negotiate the terms of their future interaction, we think it likely that they would have devised mutually beneficial rather than one-sided arrangements that would have resembled the equilibrium prediction of our model.

invested three, then the Dictator's action was the action that she indicated that she wanted to choose in the event that the Recipient chose an investment of three.

At the very end of the round, the subjects learned the payoffs that each would earn during that round if it was randomly selected for payment based on their actions or the Recipient's stated beliefs.[19]

We conducted thirteen experimental sessions with a total of 210 student subjects. We used a between-subject design, so subjects participated in only one treatment: 70 subjects participated in Expectation Damages; 70 subjects participated in No Regime; and 70 subjects participated in Control.

The experimental sessions were conducted at the Experimental Social Science Laboratory (XLab) at the University of California, Berkeley and the Experimental and Behavioral Economics Laboratory (EBEL) at the University of California, Santa Barbara. The XLab subject pool consisted of undergraduate students at UC Berkeley, and the EBEL subject pool consisted of undergraduate students at UC Santa Barbara.[20]

Subjects were assigned to visually isolated computer terminals. Beside each terminal they found paper instructions, which are reproduced in Online Appendix D. Instructions were read aloud to subjects and questions were answered individually and confidentially at the subjects' seats. The experiment was programmed and conducted using the software z-Tree (Fischbacher 2007).[21]

Each session consisted of two unpaid practice rounds followed by eight rounds, one of which was randomly chosen at the end of the experiment for payment based on the subjects' decisions in that round, and another of which was selected for payment based on the

[19]Instead of our modified dictator game in which subjects exchange promises to cooperate conditional on being selected to be the Dictator, we might instead have used a trust game with payoffs similar to those of our modified dictator game (as in Mischkowski, Stone, and Stremitzer 2019). But the trust game involves two reliance decisions: the Recipient's opt-in decision and his subsequent investment decision. So had we used such a game we wouldn't have been able to observe Dictator behavior when recipients decide not to rely on the promise at all (by opting out). And we would have only observed the effects of reliance on promising among recipients who opt in to the game, who may not be representative of the population.

[20]At UC Santa Barbara, 30 subjects participated in Expectation Damages, 30 subjects participated in No Regime, and 40 subjects participated in Control. At UC Berkeley, 40 subjects participated in Expectation Damages, 40 subjects participated in No Regime, and 30 subjects participated in Control.

[21]Sample screenshots can be downloaded from the authors' webpage.

recipient's stated beliefs at Step 3 about the likelihood that the dictator would cooperate with him for each possible investment level. Recipients were paid for stating their beliefs in accordance with Table A1 in the Appendix, which was structured to give them incentives to report their beliefs truthfully (Vanberg 2008). We selected a different round from the round selected for payment based on subjects' decisions so that subjects had no reason to change their behavior in any of the rounds to hedge their bets.

No subject interacted with the same subject more than once in the latter eight rounds. We achieved this by creating matching groups of exactly 10 subjects and having each subject play against the same subject during the two practice rounds. Thus, after the practice round, each subject was paired with eight different subjects in the matching group in a random order.

Subjects were paid an average of $11.20, ranging between $0 and $15, for their decisions, and they were paid an average of $1.88, ranging between $0 and $4.55, for stating their beliefs about the Dictator's actions. Subjects also received a fixed fee of $5 for showing up on time, $5 for completing a post-experiment survey, and an additional sum ranging between $0 and $1.50 depending on how well they did on a post-experiment cognitive reflection test.

No time limit was placed on any of the subjects' decisions. The software ensured that the experiment couldn't continue to the next round until every subject had made the decisions he was required to make in the relevant round.

In light of the complexity of the experiment, we took various steps to ensure that subjects understood the game. First, after reading the instructions aloud, subjects had to complete two preliminary questions that tested their understanding of the relationship between the players' actions and their payoffs. Subjects' answers were checked and wrong answers were corrected to make sure that they understood how to arrive at the correct answer. Second, the two practice rounds, which were designed so that subjects would experience both the role of Player A and Player B, gave subjects the opportunity to familiarize themselves with the game and the program interface before they played the game for money. Finally, in the post-experiment survey, we asked subjects whether they felt that they understood the payoff consequences of their actions. They could answer: "Yes", "No", or "Kind of." We

introduced the "Kind of" category as we thought that forcing subjects to choose between "Yes" and "No" was too crude, and we wanted to identify subjects whose understanding was only partial. We excluded subjects who answered "Kind of" and "No" from our data. This left us with 54 subjects in No Regime, 54 subjects in Control, and 57 subjects in Expectation Damages.

3.3 Hypotheses

We are now in a position to formulate the hypotheses that flow from our theoretical model. First, we predict, contrary to the predictions of the standard model, that the rate of cooperation in No Regime will exceed zero and that, anticipating cooperative behavior on the part of dictators, recipients will be willing to invest.

Hypothesis 1 *Cooperation in No Regime will exceed zero (H1.1). The rate of investment in No Regime will exceed zero (H1.2).*

Explanation. We predict that many dictators will make promises and that, of those dictators, those with high guilt parameters will cooperate regardless, while those with intermediate guilt parameters will cooperate so long as recipients invest enough. Thus, H1.2 follows so long as some recipients believe it likely that the dictator has a sufficiently high guilt parameter. H1.1 follows so long as there are in fact some dictators with sufficiently large guilt parameters.

Second, we predict that recipients will strategically overinvest in No Regime in order to psychologically lock-in the dictator, while they will not do so in Expectation Damages. We also predict that we will see less underinvestment in Expectation Damages and more efficient investment, because in Expectation Damages recipients who received a promise can give dictators a sufficient self-interested reason to cooperate simply by investing one, the efficient level of investment.

Hypothesis 2 *The incidence and magnitude of overinvestment will be higher in No Regime than in Expectation Damages (H2.1). There will be less underinvestment and more efficient*

investment in Expectation Damages than in No Regime (H2.2).

Explanation. H2.1 follows so long as some recipients hold beliefs about the dictator that make overinvestment rational. Specifically, some recipients must believe that there is a high chance the dictator has an intermediate guilt parameter that makes her willing to cooperate only if investment is greater than one. H2.2 follows from our model so long as similar numbers of promises are made in Expectation Damages as in No Regime, and some recipients aren't too confident that the dictator is a high-guilt type who will cooperate regardless of the investment level or an intermediate-guilt type who will cooperate only if investment is greater than one.

Third, we predict that dictators will become more willing to cooperate as investment increases and that recipients will anticipate this effect.

Hypothesis 3 *Recipients exhibit a higher degree of confidence that dictators will cooperate as investment increases in No Regime (H3.1). Contingent cooperation rates will increase in No Regime as investment increases (H3.2).*

Explanation. Hypothesis 3 follows from our prediction that many dictators will make promises in No Regime and so experience guilt that is increasing in a recipient's investment.

Fourth, we predict that there will be higher rates of cooperation in Expectation Damages than in No Regime.

Hypothesis 4 *The cooperation rate will be higher in Expectation Damages than in No Regime.*

Explanation. Hypothesis 4 follows from our model so long as there are some low-guilt dictators who won't cooperate in No Regime regardless of the investment level and enough promises are made in Expectation Damages compared to No Regime.

Finally, we expect that the above will all entail that Expectation Damages will do better than No Regime in terms of overall payoffs.

Hypothesis 5 *Joint payoffs will be higher under Expectation Damages than under No Regime (H5.1). Average payoff differentials will be lower under Expectation Damages than under No Regime (H5.2).*

Explanation. H5.1 and H5.2 follow if our other predictions are confirmed. This is because these predictions entail that investment decisions will be superior and rates of cooperation higher under Expectation Damages than under No Regime.

4 Results

The data from our two main treatments consist of 14 matching groups and a total of 111 subjects—54 subjects in No Regime and 57 subjects in Expectation Damages—after excluding subjects who reported that they didn't fully understand the game. Each session lasted for 8 rounds. Thus, a total of 432 decisions were made in No Regime and 456 decisions in Expectation Damages. However, the decisions made by subjects in each matching group were not independent of one another and taken together constitute just a single independent observation. Thus, while the results we report are based on data from all sessions and all rounds, our nonparametric tests are based on matching-group averages of the relevant variables.

Even though we are primarily interested in what happened when a dictator made a promise, our main results rely on aggregate data—that is, all the data including data generated when the dictator made no promise—though we sometimes also report differences focusing only on dictators who made a promise or recipients who received one. We focus on aggregate data because the dictators who made promises in No Regime may differ systematically from those who made promises in Expectation Damages. Thus, if we looked only at behavior of dictators who made a promise or recipients who received a promise, our results might be driven by selection effects. A concern about selection effects was also the motivation for running our Control treatment. Control is identical to No Regime except for the absence of a communication stage in which subjects can exchange promises. Comparing

aggregate results in Control to those in No Regime provides us with a cleaner way of assessing the effects of promising than comparing subjects who did or did not make or receive promises in No Regime, because by relying on aggregate data we eliminate concerns about selection effects.[22]

Opportunism is not as prevalent as the classical model predicts. First, consistent with much other empirical evidence, we find that the self-interested, rational actor model overestimates the importance of opportunism and so overstates the need for legal enforcement of relied-upon promises. Dictators cooperate 40.3% of the time in No Regime, despite the fact that they lack a self-interested reason to do so, and this rate of cooperation is significantly different from zero (p=0.02, Wilcoxon signed-rank test). (H1.1.) Recipients, moreover, seem to anticipate this cooperative behavior. Average investment is equal to 1.00 in No Regime, and this is significantly different from zero (p=0.02, Wilcoxon signed-rank test). (H1.2.)

A legal regime with expectation damages generates superior investment decisions. More interesting are our findings that, consistent with the existence of psychological lock-in, enforcing relied-upon promises with expectation damages induces unambiguously superior investment decisions (see Figure 3 and Table 1). Unsurprisingly, there is less underinvestment in Expectation Damages than in No Regime—32.9% compared to 55.6%—and the difference is statistically significant (p=0.047, Wilcoxon rank-sum test). (H2.2.) There is also more efficient investment in Expectation Damages—58.8% compared to 21.8%—and the difference is statistically significant (p<0.01, Wilcoxon rank-sum test). (H2.2.) Crucially, however, there is also less overinvestment in Expectation Damages—8.33% compared to 22.7%—and the difference is statistically significant (p<0.01, Wilcoxon rank-sum test). Moreover, average investment conditional on overinvesting is also lower in Expecta-

[22] We compare No Regime and Control in Section 5. The data from Control consisted of 7 matching groups and 54 subjects after excluding those who reported that they didn't fully understand the game. A total of 432 decisions were made in Control.

tion Damages—2.42 compared to 3.45—and the difference is statistically significant (p=0.02, Wilcoxon rank-sum test). These results suggest that both the magnitude and incidence of overinvestment are reduced by the introduction of a legal regime with expectation damages: more people overinvest and they overinvest more on average in the absence of a legal regime. (H2.1.) Indeed, the average *increment* of overinvestment—a measure that combines both the incidence and the magnitude of overinvestment—is also considerably lower in Expectation Damages—0.12 compared to 0.56—and the difference is again statistically significant (p<0.01, Wilcoxon rank-sum test).[23]

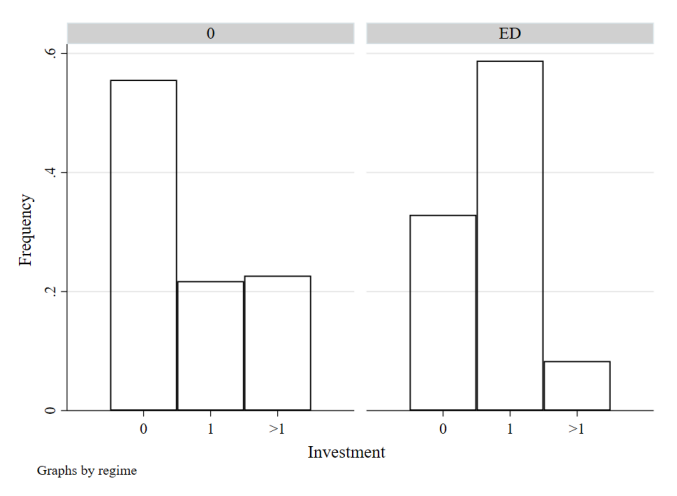


Figure 3: Incidence of Underinvestment, Efficient Investment, and Overinvestment.

Table 1: Incidence of Under-, Over-, and Efficient Investment*

	No Regime	ED	Z Stat
Efficient	0.22	0.59	$Z = 3.0^{***}$
	(0.41/216)	(0.49/228)	$(p < 0.01)$
Underinvest	0.56	0.33	$Z = 2.0^{**}$
	(0.50/216)	(0.47/228)	$(p < 0.01)$
Overinvest	0.23	0.08	$Z = 2.6^{***}$
	(0.42/216)	(0.28/228)	$(p < 0.05)$

*Pooled data from all sessions and all rounds. Standard deviation followed by N in parenthesis. The Z statistic reflects Wilcoxon Rank Tests using matching group averages.Significance at the 10%, 5%, and 1% levels is marked with *, **,*** respectively.

[23]We compute this measure by taking the mean of the increment by which actual investment exceeds the efficient level of one.

Thus, we find, consistent with our theory, that some dictators are more willing to cooperate in No Regime the more the recipient has relied on a promise, and that recipients anticipate this by overinvesting to induce more cooperation. In Expectation Damages, by contrast, there is no need to overinvest to motivate a dictator to keep her promise, because the legal regime ensures that a dictator has a sufficient self-interested reason to keep it, so long as the recipient chooses to invest at least the first-best amount of one.

Dictators' willingness to cooperate and recipients confidence increase as investment increases. We also find evidence that directly supports our theory. In line with Hypothesis 3, dictators become more willing to cooperate and recipients become more confident that dictators will cooperate when recipients overinvest—that is, invest more than one (the socially optimal level of investment). We have data to test this hypothesis because we employed the strategy method to elicit dictators' preferences for all possible investment levels and we elicited recipients' beliefs about the likelihood that the dictator will cooperate for all possible investment levels.

Our model of promisor preferences suggests that some dictators will prefer to keep their promises regardless of the investment level, others will prefer to keep their promises only when investment exceeds a certain level, while others will never cooperate regardless of the investment level. The phenomenon of psychological lock-in—overinvestment by recipients in order to lock-in a dictator—depends on the existence of this intermediate group of dictators whose preferences depend on their investment level.

If dictators have guilt parameters varying over a continuum, our model predicts that as investment increases, the number of dictators who are willing to keep their promises is non-decreasing. This is because higher investment levels induce dictators with lower guilt parameters to keep their promises while having no effect on the decisions of those with higher guilt parameters who were willing to keep their promises at lower investment levels. Thus, to test Hypothesis 3, we compare recipients' average confidence in the dictator cooperating for investment of zero in No Regime with the corresponding average for investment of six.

We also compare average contingent cooperation rates in No Regime over the same range.

In line with our hypothesis, we find that recipients' confidence increases from 1.8 to 2.2 where 0 describes the confidence of a recipient who thinks it certain that the dictator won't cooperate and 4 the confidence of a recipient who thinks it certain that the dictator will cooperate (see the line representing No Regime in Figure 4). This difference is statistically significant (p=0.03, Wilcoxon sign-rank test). (H3.1.) Importantly, the effect is not driven only by the difference between investment of zero and the efficient level of one. The difference in average confidence for investment of one (1.9) and that at investment of six is also statistically significant, albeit only at the 10% level for the Wilcoxon sign-rank test (p=0.06) and at the 5% level for the one-sided t-test (p=0.03).[24]

We also observe that the average contingent cooperation rate increases from 41% for investment of 0 to 49% for investment of 6 (see the line representing No Regime in Figure 5).[25] This difference is, however, only statistically significant at the 10% level for the one sided t-test (p=0.18, Wilcoxon sign-rank test; p=0.07, one-sided t-test).[26] (H3.2.) But focusing on the difference between cooperation rates at investment level of 1 (39%) and investment level of 6—that is, focusing specifically on the effects of overinvestment—we find statistical significance at very close to the 5% level for the Wilcoxon sign-rank test (p=0.051) and at the 5% level for the one-sided t-test (p=0.02).

[24]The one-sided t-test is appropriate where, as here, we have a directed hypothesis.

[25]Contrary to our model's prediction, we see a negative slope for average contingent cooperation in No Regime over some ranges of investment, especially when moving from $i = 2$ to $i = 3$, though the effect is not significant (p=0.13, Wilcoxon sign-rank test; p=0.14, two-sided t-test). A speculative possibility is that some dictators punish recipients for their overinvestment, perhaps because they don't like the idea, suggested by our theory, that a recipient invests beyond the productive level purely in order to make them more likely to keep a promise. We have not explicitly allowed for such backlash effects in our model. But notice that our experiments made the efficient level of reliance very transparent to subjects. Thus, it was relatively easy for our subjects to distinguish between efficient and excessive reliance. In the real world it will often be hard for a promisor to make this distinction because of imperfect information about the preferences and costs of the promisor. In this respect, our results likely overestimate a possible backlash effect.

[26]The fact that the difference in contingent cooperation is barely significant while the difference in beliefs is highly significant is likely due to the fact that beliefs are measured on a scale from 0 to 4 while the cooperation decisions are binary. This binary structure adds noise to the data. Moreover, about 88% of recipients invest 2 or less and the difference between average contingent cooperation for investment of 0 and 2 is highly significant (p<0.01,Wilcoxon sign-rank).

Introducing a legal regime increases rates of cooperation. Cooperation rates are higher in Expectation Damages than No Regime: 62.3% compared to 40.3%. The difference is statistically significant (p=0.02, Wilcoxon rank-sum test). (H4.) This result is not particularly surprising. So long as dictators make promises, thus ensuring that there is legal enforcement of relied-upon promises in Expectation Damages, the result is predicted by the standard self-interested rational actor model as well as our model in which dictators experience guilt from breaking their promises. Indeed, the difference is more stark, and again statistically significant, when we focus only on dictators who made a promise: 93.2% cooperate in Expectation Damages, while 47.5% cooperate in No Regime ($p<0.01$, Wilcoxon rank-sum test). By contrast, there is no statistically significant difference in cooperation rates among dictators who didn't make a promise: 7.7% cooperate under No Regime compared to 7.3% under Expectation Damages.

Introducing a legal regime produces crowding out of cooperation off the equilibrium path. The difference in cooperation rates between No Regime and Expectation Damages, however, masks one interesting difference in dictators' behavior that is not predicted by our model of promisor preferences and so not reflected in our hypotheses. When recipients invest zero, dictators face no pecuniary penalties if they break their promises to cooperate regardless of the regime. In No Regime, dictators are never required to compensate recipients for promise breaking, while in Expectation Damages they must only do so once the recipient has invested at least one. Thus, from the standpoint of parties' monetary incentives, Expectation Damages and No Regime are indistinguishable when recipients invest zero, and thus both the standard rational actor model and our alternative model predict no difference in cooperation rates contingent on zero investment across the regimes.

In fact, however, as Figure 5 shows, we find that the average cooperation rate contingent on zero investment is considerably lower in Expectation Damages than in No Regime: 18.9% compared to 40.7% when we look at the behavior of all dictators and 24.7% compared to 46.9% when we look only at dictators who made a promise. Both differences are statistically

significant (p=0.02 and p=0.048 respectively, Wilcoxon rank-sum test). (Among dictators who didn't make a promise, the difference is not statistically significant.)

Recipients, moreover, anticipate this effect, as shown in Figure 4. Average recipient confidence that the dictator will cooperate for zero investment is higher in No Regime than Expectation Damages: 1.8 compared to 1.1 in aggregate, and 2.1 compared to 1.4 when we look only at recipients who received a promise. Both differences are statistically significant (p=0.03 for both, Wilcoxon rank-sum test).

These results suggest that the presence of a legal regime crowds out voluntary promise keeping when the recipient fails to trigger its penalties by investing zero.[27] To capture this effect within our model of promisor preferences, we would have to assume that the legal regime reduces the guilt a dictator experiences as a result of promise breaking holding the recipient's payoffs constant (perhaps because the dictator interprets the recipient's decision to invest zero thereby failing to trigger legal penalties for promise breaking as releasing the dictator from her obligation to keep her promises). That is, we would need to assume that the unrealized possibility of legal consequences for promise breaking crowds out the dictator's internal motivation to keep her promises by reducing her guilt parameter whenever the recipient decides not to trigger those consequences by investing zero.

This crowding out seems unlikely to have a large effect on overall cooperation rates, however, because recipients can prevent it from occurring, and guarantee the optimal result, by investing one if they received a promise. Indeed, the data confirm that this crowding out largely occurs off the equilibrium path. In Expectation Damages, only 4.0% of recipients invest 0 if they received a promise (compared to 52.2% in No Regime and 89.6% in

[27] See Frey and Jegen (2000) for a survey of experimental evidence for crowding effects in economic settings.

Expectation Damages if there was no promise).

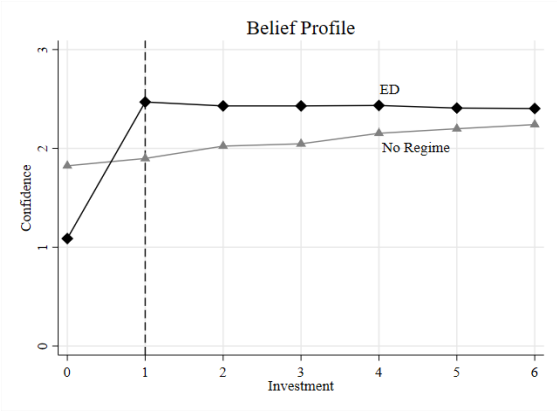


Figure 4

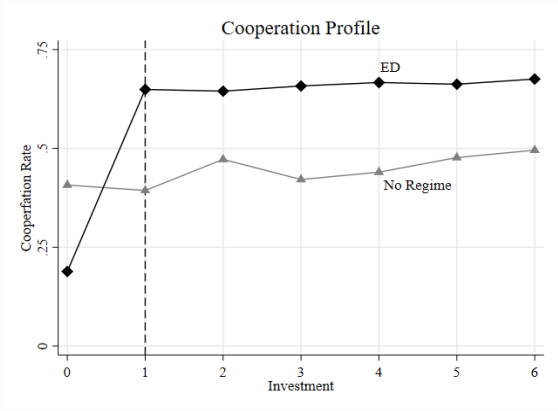


Figure 5

With the exception of this crowding out, Expectation Damages performs exactly as standard theory predicts. And our results give us no reason to expect crowding out to occur at any level of investment in a regime that enforced promises regardless of the recipient's reliance, because in such a regime a promisee would be legally protected against breach regardless of his investment decision. In short, our results suggest that a reliance-independent Expectation Damages regime would conform to the predictions of standard theory over the entire investment range. Thus, we see no reason to suppose that our main finding that legal enforcement prevents overinvestment is confined to a setting where legal enforcement is contingent on the recipient relying on a promise. Rather, it seems to be the enforcement of promise breaking with expectation damages that is driving this result.

Introducing a legal regime increases joint payoffs and reduces payoff differences. The improved investment decisions and higher cooperation rates that result from introducing the legal regime result in higher joint payoffs. Average joint payoffs are \$22.86 in Expectation Damages compared to \$21.68 in No Regime and the difference is significant (p=0.01, Wilcoxon rank-sum test). The difference is even starker when we look only at pairs of subjects where the dictator made a promise: \$23.92 compared to \$21.87 (p<0.01, Wilcoxon rank-sum test). There is no statistically significant difference in average joint

payoffs among pairs of subjects in which the dictator didn't make a promise. (H5.1.)

The introduction of a legal regime also reduces average payoff differences between dictators and recipients from \$5.90 to \$3.15, a statistically significant reduction (p=0.02, Wilcoxon rank-sum test). The difference is even more stark when we focus on pairs of subjects in which the dictator made a promise: \$5.28 compared to \$0.08 (p<0.01, Wilcoxon rank-sum test). Again, there is no statistically significant difference in average joint payoffs among pairs of subjects in which the dictator didn't make a promise.

Average payoff differences are a measure of risk resulting from opportunism. Before subjects were assigned their roles, they didn't know whether they would be assigned the role of Dictator or Recipient, and so they didn't know whether they would be in a position to exploit the other subject or instead be exploited. Our results suggest that behind this veil of ignorance, a risk-averse subject would prefer Expectation Damages to No Regime, because it reduces the risk that he will be exploited by the other ex post.[28] (H5.2.)

Subjects appeared to understand the game. As we explained above, we went to considerable lengths to try to make sure that subjects understood the game. Of course, the question remains whether they did in fact understand the game. Figure 5 shows that average contingent cooperation in Expectation Damages rises sharply at an investment level of one when a promise was made and remains high for higher investment levels. And Figure 4 shows that recipients become close to completely confident that a dictator will cooperate as soon as investment is greater or equal to one when a promise was made. Both suggest that subjects understood their payoffs in Expectation Damages. In No Regime, average contingent cooperation rates and average confidence levels rise more steadily as investment increases consistent with our posited psychological lock-in effect.

[28]This result is likely to generalize because it arises from the fact that expectation damages insure the promisee against breach. In the absence of a legal regime, the promisee cannot avail himself of this kind of insurance.

5 Discussion

Real-World Significance. Our results suggest that ordinary people are more willing to keep their promises the more they have been relied upon by promisees and that promisees anticipate this effect of reliance and overinvest accordingly. But what about promises made in high-stake transactions by sophisticated self-interested actors? What about promises made by collective entities like corporations? It is perhaps less plausible to suppose that such agents exhibit the guilt-averse preferences of the promisor in our model that give promisees strategic reasons to overinvest in reliance on promises.

There are, however, two reasons to believe that our results may be more significant than first appears. First, the agency relationships that pervade collective entities may boost the effects of psychological lock-in. This is because a corporation's agents may feel morally bound by the promises they make on behalf of the corporation. This would make it psychologically costly for them to break such promises, while the cost of keeping such promises will be borne by the corporation rather than its agents (so long as the corporation hasn't introduced an incentive scheme for its agents that forces them to internalize those costs).

Second, if the underlying driver of a promisor's guilt tracks general beliefs about moral reasons to keep promises, we should expect third-party observers to judge a promisor more harshly when she breaks a promise that has been relied upon more by the promisee. 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 social norms, we should expect that those third parties will punish the promisor more harshly the greater was the promisee's reliance.[29] Reputational concerns that track agents' moral beliefs would then make even self-interested promisors act *as if* they had guilt-averse preferences.

In Stone and Stremitzer (2019) we provide experimental evidence that confirms this hy-

[29] 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.

pothesis using a trust game with similar payoffs to those in the current experiment. Subjects were asked to imagine that they had observed various scenarios emerging from a game that was similar in crucial respects to that played by subjects in the current experiment.[30] The scenarios differed along two dimensions: whether the party making the cooperation decision (Player A) promised the other (Player B) that she would cooperate with B or merely stated an intention to do so while making explicit that she was making no promise; and B's level of investment. After being presented with 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. Subjects saw the scenarios in a randomized order, which allowed us to generate both between-subject data (subjects' responses to the first condition) and within-subject data (subjects' responses in their entirety). The results suggest that third-party observers are more willing to punish a non-cooperative player who has made a promise than one who has not, that their willingness to punish increases with B's investment, and that there is something distinctively promissory about the latter effect, because the willingness to punish A increases more with B's investment if A made a promise to cooperate. Thus, lock-in may arise even when promisors are self-interested, if promisees increase their reliance investments to make it more likely that promise breaking will be punished by a third-party.[31]

Competing Hypotheses. Our results suggest that, even in the absence of a legal regime that enforces relied-upon promises, a promisee's reliance on a promise makes the

[30]We used a trust game instead of the more complicated exchange of conditional promises we used in the current experiment, to simplify explanation of the scenario to participants. Otherwise the game was the same: one player, who may or may not have received a promise to cooperate from the other, made an investment decision prior to learning whether the other would in fact cooperate with him, and payoffs from investment and cooperation were the same as in the current experiment.

[31]Note that the purpose of this experiment was not to provide evidence that people engage in altruistic punishment, but rather to provide evidence for the claim that to the extent people do exhibit such a propensity (as the literature referenced in note 35 suggests that they do), they will exhibit a greater willingness to punish promise breaking the more the promisee relied on the promise. That the costs of punishing were purely hypothetical may well weaken conclusions that can be drawn about the magnitude of subjects' willingness to altruistically punish others in the face of real stakes, but is less likely to weaken conclusions that can be drawn about the relative strength of such a willingness as a function of our variables of interest.

promisor more likely to keep it. We have offered a theory of promisors' preferences that can account for this result. But there are other explanations that might lead to similar predictions even in the absence of any promising. In particular, the experimental literature on cooperative behavior in dictator games, ultimatum games, and public goods games suggests that cooperative behavior can arise from efficiency or equity preferences. Such preferences could cause the Dictator to cooperate as investment increases and the Recipient to overinvest in the absence of the promissory mechanism that we theorize is driving subjects' behavior.[32]

One reason to believe that our posited promissory mechanism is driving our results can be found in the existing literature on the effect of promisees' expectations on promise keeping. While the evidence as to whether a counterparty's expectations alone increase a person's willingness to cooperate is mixed,[33] there is evidence that expectations certainly do increase a promisor's willingness to keep her promises when those expectations were induced by her promise (Ederer and Stremitzer 2017; Mischkowski, Stone, and Stremitzer 2019). To the extent that a promisee's reliance on a promise to perform has a similar effect on a promisor's willingness to cooperate as a promisee's expectation of performance, our model's assumptions are supported by these findings.[34]

The comparison between No Regime and Control also provides evidence that there is something distinctively promissory about our lock-in effect. Control was identical to No Regime with the exception that participants were not able to send messages to each other. Thus, promising was entirely absent in our Control treatment, and therefore comparing Control and No Regime allows us to cleanly evaluate the influence of promising.[35] Consistent

[32]See Online Appendix B for a detailed descriptions of how efficiency or equity preferences play out in our setting.

[33]Vanberg (2008) and Ellingsen et al. (2010) find that they don't matter, while Reuben et al. (2009) and Mischkowski, Stone, and Stremitzer (2019) find that they do.

[34]Reliance and expectations have a similar flavor but there are important differences. Expectations don't have direct monetary payoff consequences. Expectations are also not as easily manipulable by the holder of those expectations. It is difficult to see how a recipient could strategically increase his expectations of cooperation to induce a dictator to cooperate, though a message by a recipient about his expections could be used to strategically overstate his expectations (and, indeed, Ellingson et al. 2010, allow subjects to send such messages).

[35]In No Regime not all dictators made promises (and therefore not all recipients received them), so we have variability in data from this treatment that might seem to allow us to test whether there is something distinctively promissory about our lock-in effect. But because promises are generated endogenously, the

with the conjecture that there is something distinctively promissory about the lock-in effect, more recipients overinvest in No Regime compared to Control: 22.7% compared to 13.9%. This effect is statistically significant (Wilcoxon rank sum, p=0.08; one-sided t-test p=0.03). For a more detailed comparison of the two treatments that makes use of the belief and contingent cooperation data, see Online Appendix B.

Finally, Stone and Stremitzer (2019) find not only that third parties are more willing to punish, at some cost to themselves, non-cooperative promisors when their promises were relied upon by the promisee, but also that the effect of reliance on the willingness of third-parties to punish non-cooperative action is greater when that non-cooperation breaches a promise to cooperate as opposed to simply running counter to a prior statement of intent. To the extent that such altruistic punishment is driven by the same forces that drive promisors to keep their promises in the first place, this serves as further evidence that lock-in has promissory underpinnings.

Incidence of Promising under No Regime and Expectation Damages. For reasons explained in Section 3.1 and Online Appendix A, our model predicts that subjects will always make promises in Expectation Damages, while they will only do so in No Regime if they believe that the subject they are paired with has a sufficiently high guilt parameter. Therefore, we should expect no more promising under No Regime than under Expectation Damages.

Contrary to this prediction, Expectation Damages causes a reduction in the numbers of promises that get made: 81.7% of subjects make promises in No Regime compared to 63.6% in Expectation Damages. The difference is significant, although only at 10% (p=0.06, Wilcoxon rank-sum test). This suggests that factors outside our model are driving promise making behavior. This is unsurprising given that anticipating the full game at the beginning of the game requires a lot of sophistication on the part of subjects. If subjects myopically focus on the material costs of promising without weighing those costs against the potential

unobserved characteristics of dictators who promise and those who don't may systematically differ, which prevents us from cleanly testing this hypothesis using these data.

benefits, rates of promising will be lower under Expectation Damages in line with what we find.

The fact that we observe this difference between the regimes, however, doesn't undermine our central results. Expectation Damages produces higher cooperation rates in aggregate than No Regime despite the lower incidence of promising. And the average increment of overinvestment when dictators made a promise is significantly higher in No Regime than in Expectation Damages: 0.57 compared to 0.16 (p=0.02, Wilcoxon rank-sum test), which suggests that psychological lock-in is driving the greater average overinvestment in No Regime. Moreover, as explained in the previous subsection, the comparison between Control and No Regime suggests that promissory lock-in explains our results.

Strategy Method and Demand Effects. It is possible that some features of our design may have conveyed to subjects that dictators were expected to alter their responses in response to recipients' investment decisions, which might have induced subjects to behave in a manner conforming to our hypothesis. In particular, because we employed the strategy method to determine dictators choices, we elicited dictators' preferences for all possible investment levels, including those not actually chosen by recipients. We also asked recipients to predict what the dictator would do for all possible investment levels. And each subject experienced the roles of both dictator and recipient over the course of the experiment.[36]

Yet nothing about our design suggested to subjects that they were expected to respond to the investment variable in a manner conforming to our hypotheses.[37] Thus, we think that

[36]It is inherently difficult to simulate the real world salience of a variable in a lab experiment. Our design may have made a promisee's reliance more salient to subjects than it would have been in the real world. But it is also likely that observable reliance by the promisee will be highly salient to the promisor in the real world. Indeed, our theory suggests that the promisee has a strong incentive to make any reliance known to the promisor. And other experimental work has shown that a promisee's expectations of performance are important determinants of promise keeping (Charness and Dufwenberg 2006; Ederer and Stremitzer 2017; Mischkowski, Stone, and Stremitzer 2019). A promisee's reliance and expectations are closely related phenomena insofar as the promisee's willingness is likely depend on his expectations of performance.

[37]Our Control treatment made the investment level salient to subjects in exactly the same ways as our No Regime treatment did. Despite this similarity, the comparison between the two suggests that there is something distinctively promissory about the lock-in effect as explained above. Also of note is Brandts and Charness's (2011) survey of experimental studies that employ both strategy and direct-response methods (where subjects respond only to realized values of variables of interest), which found that "there are signifi-

our results still tell us something about the *existence* of a real-world effect, even though we cannot make inferences about its magnitude.[38]

6 Implications for Legal Theory

Implications for Enforcement Theory. We find that in the absence of a legal regime promisors are more likely to keep promises that have been relied upon. This result, in conjunction with Stone and Stremitzer's (2019) finding that third parties also punish relied upon promises more harshly, suggests that people believe that it is morally worse to break a promise the more the promisee relied upon it. Furthermore, promisees anticipate the lock-in effect of reliance on promisors and so overinvest to induce their promisors to keep their promises.

To the extent that one believes that law tracks prevailing moral intuitions, our results help to explain the emergence of legal doctrines that create liability for relied-upon promises. The best example of such a doctrine in the common law is promissory estoppel, which gives significance to a promisee's reliance by treating it as a ground of enforcement.

Our findings can also be viewed as adding to the economic theory of legal enforcement. The standard view is that legal enforcement serves only to mitigate underinvestment that arises when a promisee fears that the promisor won't perform. Our findings point to a surprising additional benefit of legal enforcement—a reduction of overinvestment that arises from a promisee's incentive to psychologically lock-in a promisor. Thus, legal enforcement may actually reduce overinvestment at the same time as it also reduces underinvestment.

Of course, the recipient of a promise that is supported by consideration but unenforceable due to lack of definiteness or failure to satisfy the Statute of Frauds could ensure enforcement of the promise by persuading the promisor to make his promise more definite or ensuring that the Statute of Frauds is satisfied rather than by investing to psychologically lock-in the

cantly more studies that find no difference across elicitation methods than studies that find a difference" (p. 387).

[38]It is generally true of lab experiments that while they can point to the existence of effects, ineliminable concerns about external validity prevent inferences being drawn about effect sizes. Reliable results about effect sizes require field experiments.

promisor. But there are often significant transaction costs of making an agreement more definite and formal: formalizing details takes time; the parties' interests may diverge in the interim; and insisting on formality or further detail may be viewed by the promisor as a signal of the promisee's future litigiousness or some other negative characteristic, which could unravel the agreement.[39] By contrast, reliance on a promise is a unilateral act that is less likely to be viewed as an aggressive act at least if such reliance is reasonably foreseeable.

Enforcement Induced Overinvestment. It might seem that Shavell's claim that legal enforcement of promises with expectation damages produces overinvestment undermines the significance of our claim that such enforcement mitigates an overinvestment problem (Shavell 1980, 1984).[40] But Shavell's prediction is premised on there being a positive probability that performance will turn out to be inefficient ex post. This was not a feature of our experimental design, and in many real-world situations performance is always, or nearly always, the efficient choice. Shavell also assumes that the performance decision is binary. But when the performance variable is instead continuous, parties who know that their agreement will be legally enforced with expectation damages can set the performance term to ensure that the promisee has no incentive to over- or underinvest (Edlin and Reichelstein 1996).[41] Finally, as Cooter and Eisenberg (1985) have pointed out, even when the performance variable is binary, the promisee's incentive to overinvest can be eliminated by limiting damages to the promisee's expectation had she not overinvested—something courts would arguably have the discretion to do in the promissory estoppel context where remedies may be "limited as justice requires"(Restatement (2d) of Contracts § 90, 1981).

[39]In the spirit of this suggestion, Spier (1992) argues that asymmetric information about a contracting party's type can drive contractual incompleteness.

[40]Intuitively, a self-interested promisee invests the amount that would be optimal if performance were always efficient, because the promisee is fully compensated for breach, and so will fail to internalize the likelihood that performance will be inefficent when making his investment decision.

[41]Intuitively, the promisee would have an incentive to underinvest if the agreed level of performance were zero (equivalent to no enforcement) and would have an incentive to overinvest if the agreed level of performance exceeded a certain level. The parties can balance the two effects by agreeing to an intermediate level of performance.

Mitigating Underinvestment. When a promisee can motivate the promisor to perform in the absence of legal enforcement simply by investing at the efficient level, the forces of psychological lock-in will mitigate underinvestment rather than producing overinvestment. Whether an enforcement regime's potential to reduce overinvestment is more or less significant than its potential to reduce underinvestment depends on the distribution of types of contracting parties and their beliefs about their contracting partners—specifically their guilt parameters. Notice, however, that enforcement should be able simultaneously to reduce the underinvestment of some and the overinvestment of others. Suppose, for example, that there are three agents: one who invests zero in the absence of legal enforcement because he believes that his partner has a low guilt parameter; another who invests one because he believes that his partner has a large guilt parameter; and another who invests two because he believes that his partner's guilt parameter is in the intermediate range. Our findings suggest that legal enforcement can correct both problems simultaneously by encouraging the underinvesting agent to invest more and the overinvesting agent to invest less.

7 Conclusion

We show that reliance on promises matters, because many promisors are more likely to keep promises that have been relied upon. Promisees, moreover, anticipate this, and accordingly some overinvest in order to motivate promisors to keep their promises. Thus, we show that there is a heretofore unappreciated benefit of legally enforcing relied-upon promises with expectation damages: legal enforcement reduces the overinvestment that results when promisees invest in order to psychologically lock in promisors.

Our results also have a specific doctrinal implication. While the doctrine of promissory estoppel tracks our subjects' moral intuitions by according legal significance to the promisee's reliance, the unexpected benefit of legal enforcement that we identify—the reduction in overinvestment—is less likely to be realized when a showing of actual reliance is made a condition for legal enforcement. This is because the tendency to overinvest is eliminated only if the promise is enforced, but if a promisee must invest in order to ensure the promise

is enforced, then he may have to overinvest just to get the benefit of the legal regime. In *Maryland National Bank v. United Jewish Appeals Federation* (407 A.2d 1130 [Md. 1979]), for example, the court found that the promisee charity did not have a viable promissory estoppel claim against the promisor's estate (the promisor having died before he had honored his promise to the charity), because it had engaged in prudent budgeting, failing to pledge any unpaid portions of the promised donation to its charitable projects, with the result that it had not in fact detrimentally relied on the promise. Had it instead pledged some of the promised funds to specific projects, even imprudently or prematurely, it might well have prevailed. Accordingly, it will often make sense for lawyers to advise their clients to engage in detrimental reliance after they have received a promise so that they can invoke promissory estoppel in the event that the promise is broken. This suggests that enforcement pursuant to the doctrine of promissory estoppel may create its own incentives for promisees to overinvest (e.g., the Teacher's Manual of the Contracts Casebook by Tracey E. George and Russell Korobkin). Can these incentives to overinvest be eliminated by getting rid of promissory estoppel? Our findings suggest not: incentives to overinvest exist and might even be higher in the absence of legal enforcement. Therefore, reducing the scope of promissory estoppel will not solve and could well aggravate the overinvestment problem.

The solution lies instead in reducing the amount of actual reliance that is required to trigger enforcement of a promise or eliminating the requirement that the promisee actually rely on the promise entirely. Enforcing more promises eliminates a promisee's incentive to overinvest in order to psychologically lock-in the promisor, and reducing the amount of reliance required to trigger enforcement makes it less likely that a promisee will have to overinvest in order to get the protection of the doctrine.

In line with this prescription, the doctrine appears to be moving away from requiring a showing of reliance. The Second Restatement of Contracts dropped the first Restatement's requirement that the reliance on the promise be of a "definite and substantial character" (Restatement of Contracts § 90, 1932). And in the light of their survey of the case law, Yorio and Thel (1991, pp. 156-57) argue that courts don't in fact do a searching inquiry for

actual reliance in promissory estoppel cases, focusing instead on the extent to which reliance was reasonably foreseeable to the promisor as a marker of the promisor's seriousness when he made the promise. Farber and Matheson (1985, p. 910) likewise conclude that "the requirement of an identifiable detriment no longer defines the boundary of enforceability" when courts are applying the doctrine of promissory estoppel.

Thus, our results point to a novel reason why courts should be more willing to recognize promissory estoppel claims and less insistent on requiring actual reliance. But our results likely have wider applicability. We designed our Expectation Damages treatment to mimic the operation of a promissory estoppel regime. Our main result, however, would likely generalize to a legal regime that enforced promises in the absence of any reliance. In this sense our results shed light on the benefits of contract enforcement more generally.

References

Aghion, Philippe, Mathias Dewatripont, and Patrick Rey. 1994. Renegotiation Design with Unverifiable Information. *Econometrica* 62: 257–282.

Battigalli, Pierpaolo, and Martin Dufwenberg. 2007. Guilt in Games. *American Economic Review* 97(2): 170–176.

Battigalli, Pierpaolo, and Martin Dufwenberg. 2009. Dynamic Psychological Games. *Journal of Economic Theory* 144(1): 1–35.

Bicchieri, Cristina, and Azi Lev-On. 2007. Computer-Mediated Communication and Cooperation in Social Dilemmas: An Experimental Analysis. *Politics, Philosophy, and Economics* 6(2): 139–168.

Brandts, Jordi, and Gary Charness. 2011. The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics* 14(3): 375–398.

Braver, Sanford L. 1995. Social Contracts and the Provision of Public Goods. Pp. 69–86 in *Social Dilemmas: Perspectives on Individuals and Groups*, edited by David Schroeder. New York: Prager.

Charness, Gary, and Martin Dufwenberg. 2006. Promises and Partnership. *Econometrica* 74: 1579–1601.

Charness, Gary, and Martin Dufwenberg. 2010. Bare Promises: An Experiment. *Economics Letters* 107: 281–83.

Che, Yeon-Koo, and Donald B. Hausch. 1999. Cooperative Investments and the Value of Contracting. *American Economic Review* 89: 125–147.

Che, Yeon-Koo, and Tai-Yeong Chung. 1999. Contract Damages and Cooperative Investments. *RAND Journal of Economics* 30: 84–105.

Chung, Tai-Yeong. 1991. Incomplete Contracts, Specific Investments, and Risk Sharing. *Review of Economic Studies* 58: 1031–1042.

Cooter, Robert, and Melvin Aron Eisnberg. 1985. Damages for breach of contract. *Calif. L. Rev.* 73: 1432.

Dufwenberg, Martin, and Uri Gneezy. 2000. Measuring Beliefs in and Experimental Lost Wallet Game. *Games and Economics Behavior* 30: 163–182.

Ederer, Florian, and Alexander Stremitzer. 2017. Promises and Expectations. *Games and Economic Behavior* 106: 161–178.

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

Edlin, Aaron S., and Stefan Reichelstein. 1996. Holdups, Standard Breach Remedies, and Optimal Investment. *American Economic Review* 86: 478–501.

Ellingsen, Tore, and Magnus Johannesson. 2004. Promises, Threats and Fairness. *Economic Journal* 114: 397–420.

Ellingsen, Tore, Magnus Johannesson, Sigve Tjøtta, and Gaute Torsvik. 2010. Testing Guilt Aversion. *Games and Economic Behavior* 68: 95–107.

Farber, Daniel A., and John H. Matheson. 1985. Beyond Promissory Estoppel: Contract Law and the "Invisible Handshake.". *U. Chi. L. Rev.* 52: 703.

Farnsworth, E. Allen. 2004. *Contracts 4th ed.* Aspen.

Fehr, Ernst, and Urs Fischbacher. 2004. Third-party punishment and social norms. *Evolution and human behavior* 25: 63–87.

Fischbacher, Urs. 2007. Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics* 10(2): 171–178.

Frey, Bruno S., and Reto Jegen. 2001. Motivation Crowding Theory. *Journal of Economic Surveys* 15(5): 589–611.

Gintis, Herbert, Samuel Bowles, Robert Boyd, and Ernst Fehr. 2005. 1 Moral Sentiments and Material Interests: Origins, Evidence, and Consequences. *Moral sentiments and material interests* 1.

Goetz, C.J.;, and R.E. Scott. 1980. Enforcing Promises: An Examination of the Basis of Contract. *Yale Law Journal* 89(7): 1261–1322.

Grossman, Sanford J., and Oliver D. Hart. 1986. The Costs and Benefits of Ownership - A Theory of Vertical and Lateral Integration. *Journal of Political Economy* 94: 691–719.

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

Hart, Oliver, and John Moore. 1988. Incomplete Contracts and Renegotiation. *Econometrica* 56: 755–785.

Holmes, Eric Mills. 1996. Restatement of Promissory Estoppel. *Willamette L. Rev.* 32: 263.

Ismayilov, Huseyn, and Johannes J. M. Potters. 2012. Promises as Commitments. *Tilburg University, Center for Economic Research, Discussion Paper* 2012-064.

Kerr, Norbert L., and Cynthia M. Kaufman-Gilliland. 1994. Communication, Commitment and Cooperation in Social Dilemma. *Journal of Personality and Social Psychology* 66(3): 513–529.

Knapp, Charles L. 1998. Rescuing Reliance: The Perils of Promissory Estoppel. *Hastings Law Journal* 49: 1191.

Mischkowski, Dorothee, Rebecca Stone, and Alexander Stremitzer. 2019. Promises, Expectations, and Social Cooperation. *Journal of Law and Economics* forthcoming.

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

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

Ostrom, Elinor, James Walker, and Roy Gardner. 1992. Covenants With and Without a Sword: Self-Governance Is Possible. *American Political Science Review* 86: 404–417.

Rogerson, William P. 1984. Efficient Reliance and Damage Measures for Breach of Contract. *RAND Journal of Economics* 15: 37–53.

Sally, David. 1995. Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992. *Rationality and Society* 7(1): 58–92.

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

Scott, Robert E. 2007. Hoffman v. Red Owl Stores and the Myth of Precontractual Reliance. *Ohio St. LJ* 68: 71.

Selten, Reinhard. 1967. Die Strategiemethode zur Erforschung des eingeschraenkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. Pp. 136–168 in *Beitraege zur experimentellen Wirtschaftsforschung*, edited by H. Sauermann. Tuebingen: Mohr.

Shavell, Steven. 1980. Damage Measures for Breach of Contract. *Bell Journal of Economics* 11: 466–490.

Shavell, Steven. 1984. The Design of Contracts and Remedies for Breach. *The Quarterly Journal of Economics* 99: 121–148.

Sloof, Randolph, Edwin Leuven, Hessel Oesterbeck, and Joop Soonemans. 2003. An Experimental Comparison of Reliance Levels under Alternative Breach Remedies. *RAND Journal of Economics* 26(3): 205–222.

Sloof, Randolph, Hessel Oesterbeck, Arno Riedl, and Joop Soonemans. 2006. Breach remedies, reliance and renegotiation. *International Review of Law and Economics* 34): 263–296.

Spier, Kathryn. 1992. Incomplete Contracts and Signalling. *RAND Journal of Economics* 23(3): 432–43.

Stone, Rebecca, and Alexander Stremitzer. 2019. Norm-Based Enforcement of Promises. *UCLA School of Law, Public Law Research Paper* No. 19-04.

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

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

Wilkinson-Ryan, Tess, and David A. Hoffman. 2015. The Common Sense of Contract Formation. *Stanford Law Review* 67: 1269–1301.

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

Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism.* New York: Free Press.

Yorio, Edward, and Steve Thel. 1985. Promissory Basis of Section 90. *Yale L. J.* 101: 111.

APPENDIX

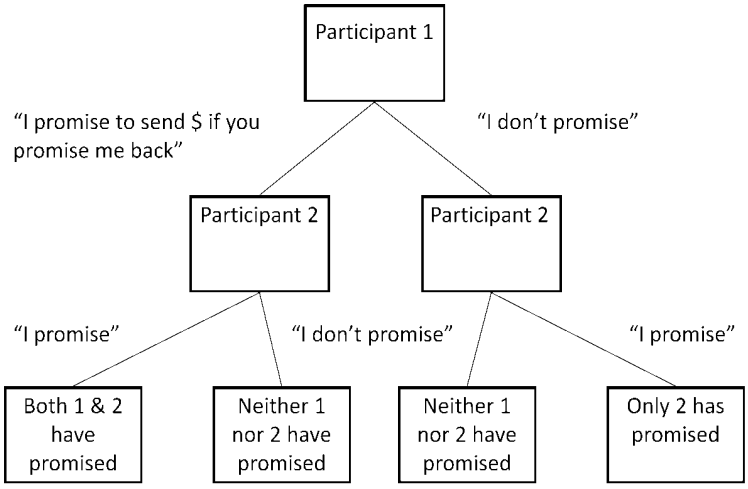


Figure A1: Design of the Communication Stage

	Player A will certainly send Player B money	Player A will probably send Player B money	There is a 50-50 chance that Player A sends Player B money	Player A probably will not send Player B money	Player A certainly will not send Player B money
Player B's earnings if Player A decides to send him/her money	$0.65	$0.60	$0.50	$0.35	$0.15
Player B's earnings if Player A decides not send Player B money	$0.15	$0.35	$0.50	$0.60	$0.65

Table A1: How Player B's Payoffs from Stated Beliefs Were Determined

ONLINE APPENDIX A: Theoretical Model

Here we set out and solve the full theoretical model that we use to generate the hypotheses that we test using the data from our experiment. We contrast the predictions derived from a model in which the promisor exhibits certain "guilt-averse" preferences with those of classical economic theory.

Modified Dictator Game

Two agents, the Dictator and the Recipient, play a game in which the Dictator must decide whether or not to cooperate with the Recipient after the Recipient makes an investment decision that only affects the Recipient's payoffs. The Recipient then chooses an investment level $i \in \left[0, \overline{i}\right]$. Finally, after observing the Recipient's choice, the Dictator chooses an action $a \in \{0, 1\}$ where $a = 1$ denotes the cooperative action.

Before learning their roles in this game, the two agents send one another messages in accordance with the procedure described in Step 1 in Section 3.2. The upshot is that they may end up having made a promise to cooperate with the other in the event that they are assigned the role of Dictator in the subgame that follows. Thus, once the modified dictator game begins, the Dictator either has or has not made a promise to cooperate with the Recipient, $p \in \{0, 1\}$. This promise has no effect on either agent's material payoff in the absence of legal enforcement.

Assumption 1. In the absence of a legal regime, the Recipient's material payoff $\pi_R(a, i)$ depends on both the investment level and the Dictator's action, while the Dictator's material payoff $\pi_D(a)$ depends only on her action. Cooperation by the Dictator increases joint material payoffs $W(a, i)$, but reduces the Dictator's material payoff, and so increases the Recipient's payoff. That is, for all i:

$$W(1, i) = \pi_R(1, i) + \pi_D(1) > \pi_R(0, i) + \pi_D(0) = W(0, i),$$

and

$$\pi_D(1) < \pi_D(0), \tag{9}$$

and therefore

$$\pi_R(1, i) > \pi_R(0, i). \tag{10}$$

Assumption 2. In the absence of a legal regime, investment affects only the Recipient's payoff. This means that the Dictator has no reciprocity-based reason to cooperate with a Recipient who chooses positive investment, and so we abstract from such motives when we model the Dictator's preferences below. Investment always reduces the Recipient's material payoff if the Dictator doesn't cooperate. But it first increases and then decreases the Recipient's payoff if the Dictator cooperates, and decreases the Recipient's payoff at a faster rate if the Dictator doesn't cooperate. Thus:

$$\pi'_R(0, i) < 0, \tag{11}$$

while

$$\pi'_R(1, i) > 0 \text{ if } i \leq e, \tag{12}$$

and

$$\pi'_R(0, i) < \pi'_R(1, i) < 0 \text{ if } i > e, \tag{13}$$

where $e \in \left(0, \bar{i}\right)$. The payoff function thus captures a situation in which there are decreasing marginal returns to investment. Marginal returns decrease discretely once investment exceeds e such that after that point investment ceases to to be productive. We assume this piecewise linear form in order to keep things simple. It follows from this assumption about the Recipient's payoff function that the extent to which cooperation increases the Recipient's material payoff is increasing in i. That is:

$$\pi'_R(1, i) - \pi'_R(0, i) > 0. \tag{14}$$

As we show in the next section, this ensures that the Dictator's preference to keep a promise is, all else equal, monotonically increasing in the Recipient's investment.

Specification of the payoff function. While many of our results hold for general payoff functions satisfying Assumptions 1 and 2, we will sometimes use specifications of the payoff functions that resemble those faced by subjects in our experiment.[42] Consistent with Assumption 1 we assume that the Dictator's material payoff is given by:

$$\pi_D(a, i) = \begin{cases} 15 & \text{if } a = 0 \\ 12 & \text{if } a = 1 \end{cases}, \tag{15}$$

and the Recipient's material payoff is given by:

$$\pi_R(a, i) = \begin{cases} 6 - i & \text{if } a = 0 \\ 12.25 - 0.25\,|1 - i| & \text{if } a = 1 \end{cases}, \tag{16}$$

where he can choose an investment level from $i \in [0, 6]$. Notice that, consistent with Assumption 2, investment always reduces the Recipient's material payoff if the Dictator doesn't cooperate. But it first increases (for $i \leq e = 1$) and then decreases the Recipient's payoff (for $i > e = 1$) if the Dictator cooperates, and it decreases the Recipient's payoff at a slower rate if the Dictator cooperates.

Preferences

Because the Recipient makes no promises, we suppose that he cares only about his material payoff. The Dictator, by contrast, cares both about her material payoff and about keeping promises that have been relied upon by the Recipient.

Assumption 3. The Recipient's overall utility u_R is simply given by his own material payoff:

$$u_R(a, i) = \pi_R(a, i).$$

Assumption 4. The Dictator's utility function u_D is determined in part by her material payoff. But it also reflects her desire not to let the Recipient down by breaking a promise to cooperate with the Recipient—a desire that is intensified the more the Recipient will lose should the Dictator break her promise. We therefore posit that, if the Dictator breaks a promise, her utility is reduced by an amount that depends on the reduction of the Recipient's material payoff resulting from the broken promise.[43] In line with the economics literature on

[42] The only difference is that, in our experiment, the Recipient chooses from a discrete choice set: $i \in \{0, 1, 2, 3, 4, 5, 6\}$.

[43] Because our focus is on the effects of reliance on promising, we abstract from other considerations that might drive agents to keep their promises like a desire not to disappoint promisees' expectations or a simple desire to do as one promised (Mischkowski, Stone, and Stremitzer, 2016).

promising, we refer to the latter component of the utility as the Dictator's "guilt" (Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007, 2009; Ederer and Stremitzer, 2017). Formally, the Dictator's utility is given by:

$$u_D(a,p) = \pi_D(a) - p\gamma g\left(\pi_R(1,i) - \pi_R(a,i)\right), \tag{17}$$

where $\gamma \geq 0$ is a parameter that represents the degree of the Dictator's guilt and $g(x)$ is the guilt function with $g(0) = 0$ and $g'(x) > 0$. Guilt is therefore zero whenever the Dictator does not promise, $p = 0$, and whenever she cooperates, $a = 1$, so $u_D(a,0) = \pi_D(a)$ and $u_D(1,1) = \pi_D(1)$. Expression (14) implies that guilt from not keeping a promise increases in investment, so $\frac{dg}{di} > 0$ if $a = 0$.

Complete Information Equilibrium of Final Subgame without Legal Enforcement

We now solve the game by backwards induction for the equilibrium of this game without legal enforcement of promises $\left(\widehat{a},\widehat{i}\right)$. We begin by assuming that the Dictator's preferences—in particular, her guilt parameter—are known to the Recipient. Later we will relax that assumption. We start with the subgame of the game that follows the communication stage and so for the moment we take the Dictator's decision whether to make a promise as exogenous.

If the Dictator experiences no guilt when she breaks a promise, $\gamma = 0$, and so cares only about her material payoff, social welfare is maximized if the Dictator cooperates and the Recipient invests e. Proposition 1 summarizes. It follows immediately from Assumptions 1 and 2 and so is stated without proof.

Proposition 1 *The social welfare maximizing outcome is given by* $(a^*, i^*) = (1, e)$. *Maximized social welfare is given by* $W^* = W(1,e) = \pi_D(1) + \pi_R(1,e)$. *For our particular specification of the payoff functions,* $(a^*, i^*) = (1,1)$ *and* $W^* = 24.25$.

In equilibrium, however, such a Dictator never cooperates and the Recipient, anticipating this, invests zero. Thus, the equilibrium falls short of the social optimum. Proposition 2 summarizes.

Proposition 2 *If both parties only care about material payoffs, then, regardless of whether the Dictator made a promise to cooperate with the Recipient, the Recipient will invest zero, and the Dictator won't cooperate:* $\left(\widehat{a},\widehat{i}\right) = (0,0)$. *Social welfare is given by* $\hat{W} = W(0,0) = \pi_D(0) + \pi_R(0,0)$. *For our particular specification of the payoff functions,* $\hat{W} = 21$.

Proof. If the Dictator chooses not to cooperate, it is optimal for the Recipient to choose zero investment given expression (11). If the Dictator chooses to cooperate, then expressions (12) and (13) imply that the recipient will choose e. But expression (9) implies that the Dictator's dominant strategy is not to cooperate. ■

The social optimum is unaffected by the introduction of "guilt" for breaking a promise into the Dictator's utility function, regardless of whether we view this guilt as a component of the Dictator's welfare (see expression 17 above). This is because aggregate material payoffs are maximized when the Dictator cooperates and the Recipient invests e, and guilt is zero when the Dictator cooperates. Thus, the social optimum is the same regardless of whether we include the Dictator's guilt in the social welfare function. Proposition 3 summarizes. It is stated without proof, since it follows immediately from Assumption 4 and Proposition 1.

Proposition 3 *The social welfare maximizing outcome is given by* $(a^*, i^*) = (1, e)$ *regardless of whether or not we include "guilt" in the social welfare function and regardless of the value of* γ*. Maximized social welfare is given by* $W^* = W(1, e)$*.*

However, the equilibrium may differ from the no-guilt equilibrium. If a Dictator made no promise, $p = 0$, or the Dictator's sensitivity to guilt is zero, $\gamma = 0$, then since $\pi_D(1) < \pi_D(0)$ the Dictator will not cooperate at the final stage. Anticipating this, the Recipient will invest zero, because $\pi_R(0, i)$ is strictly decreasing in i. If, by contrast, the Dictator made a promise, then the Dictator will cooperate whenever

$$\pi_D(1) \geq \pi_D(0) - \gamma g\left(\pi_R(1, i) - \pi_R(0, i)\right),$$

that is, whenever the sensitivity to guilt, γ, exceeds a critical value, $\gamma^c(i)$, which depends on the Recipient's investment level:

$$\Leftrightarrow \gamma \geq \gamma^c(i) = \frac{\pi_D(0) - \pi_D(1)}{g\left(\pi_R(1, i) - \pi_R(0, i)\right)}. \tag{18}$$

Thus, promising now matters, as breaking a promise triggers guilt. In addition, the Dictator's utility is no longer independent of the Recipient's investment choice. The guilt that she experiences when she breaks a promise increases in investment. And so, if the Dictator made a promise she will be more inclined to keep her promise if investment was high than if it was low. This, in turn, might invite the Recipient to strategically invest in order to psychologically lock the Dictator in to keeping her promise. We know from expression (14) that increasing investment i increases $\pi_R(1, i) - \pi_R(0, i)$ and so, given (18), reduces $\gamma^c(i)$, which may thereby compel the Dictator to keep her promise depending on the value of γ. Proposition 4 shows how the equilibrium depends on the guilt parameter.

Proposition 4 *In the absence of a legal regime, the equilibrium is as follows:*

$$(\hat{a}, \hat{\imath}) = \begin{cases} (0, 0) & \text{if } p = 0, \text{ or } p = 1 \text{ and } \gamma < \gamma^c(\bar{\imath}) \\ (1, i^c) & \text{if } p = 1 \text{ and } \gamma \in [\gamma^c(\bar{\imath}), \gamma^c(e)) \\ (1, e) & \text{if } p = 1 \text{ and } \gamma \geq \gamma^c(e) \end{cases}$$

where $i^c(\gamma) \in (e, \bar{\imath}]$ *is decreasing in* $\gamma \in [\gamma^c(\bar{\imath}), \gamma^c(e))$*.*

Proof. If there was no promise, the equilibrium follows immediately from inserting $p = 0$ into the the utility function (17) and the proof of Proposition 2. If there was a promise, the Recipient's payoff is given by $\pi_R(1, i)$ if $\gamma \geq \gamma^c(i)$ and $\pi_R(0, i)$ otherwise. Expressions (14) and (18) imply that $g(\cdot)$ is increasing in i. Thus, $\gamma^c(i)$ is decreasing in i. If $\gamma < \gamma^c(\bar{\imath})$, then no investment level in the Recipient's choice set will induce the Dictator to cooperate, and therefore $\left(\widehat{a}, \widehat{\imath}\right) = (0, 0)$. But if $\gamma \geq \gamma^c(\bar{\imath})$, then there exists a critical investment level $i^c \leq \bar{\imath}$ such that whenever $i \geq i^c$, a Dictator for whom $\gamma \geq \gamma^c(i^c) \geq \gamma^c(\bar{\imath})$ will cooperate. Thus, if $\gamma \geq \gamma^c(e)$, it follows that $i^c \leq e$. Then, because $\pi_R(1, i)$ is maximized when $i = e$ and $\pi_R(1, e) > \pi_R(1, 0) > \pi_R(0, 0)$ by expression (10), the Recipient will always choose e, and the Dictator will cooperate: $\left(\widehat{a}, \widehat{\imath}\right) = (1, e)$. If, by contrast, $\gamma^c(\bar{\imath}) \leq \gamma < \gamma^c(e)$, then $i^c \in (e, \bar{\imath}]$. As $\max_{i \in [i^c, \bar{\imath}]} \pi_R(1, i) = \pi_R(1, i^c)$, the Recipient will choose $\widehat{\imath} = i^c$ and so $\left(\widehat{a}, \widehat{\imath}\right) = (1, i^c)$

if $\pi_R(1, i^c) > \pi_R(0, 0)$, and $(\widehat{a}, \widehat{i}) = (0, 0)$ otherwise. ■

Complete Information Equilibrium of Final Subgame with Legal Enforcement

Introducing a legal regime with expectation damages that enforces relied-upon promises forces the Dictator to make a payment to the Recipient in the event she breaks a promise and the Recipient relied on the promise by choosing a positive investment level, as described in the following assumption.

Assumption 5. If the Dictator made a promise, and the Recipient made a positive investment in reliance of the promise, the Dictator must make a payment $l(i)$ to the Recipient that is sufficient to ensure that the Recipient is as well off in material terms as if the Dictator had kept her promise. In other words, if she breaks her promise, she must pay the Recipient expectation damages:

$$l(i) = \begin{cases} 0 & \text{if } i = 0 \\ \pi_R(1, i) - \pi_R(0, i) & \text{if } i > 0 \end{cases}.$$

If the Dictator breaks her promise by choosing not to cooperate, the Dictator's payoffs are therefore:

$$\pi_D^L(0, i) = \pi_D(0) - l(i) = \begin{cases} \pi_D(0) & \text{if } i = 0 \\ \pi_D(0) - l(i) & \text{if } i > 0 \end{cases}, \tag{19}$$

and the Recipient's payoffs are:

$$\pi_R^L(0, i) = \pi_R(0, i) + l(i) = \begin{cases} \pi_R(0, 0) & \text{if } i = 0 \\ \pi_R(1, i) & \text{if } i > 0 \end{cases}.$$

Payoffs are unchanged if the Dictator cooperates. Thus:

$$\pi_D^L(1, i) = \pi_D(1), \tag{20}$$

and

$$\pi_R^L(1, i) = \pi_R(1, i).$$

Assumption 5 implies that $\pi_R^L(0, i) = \pi_R^L(1, i)$ so long as $i > 0$. That is, given that the Recipient chooses a positive investment level, he will always receive $\pi_R(1, i)$, regardless of whether the Dictator cooperates or not. Thus, it is easy to see that the investment level that maximizes the Recipient's payoff is the socially optimal investment level, $i = 1$.[44] Because the Recipient's payoff is unaffected by the Dictator's action if he relied on the Dictator's promise, the Dictator experiences no guilt if she breaks a promise since the Recipient's payoffs are unaffected by the Dictator's decision: $u_D^L(a, 1, i) = \pi_D^L(a, i)$ so long as $i > 0$. Thus, the equilibrium $\left(\widehat{a}_L, \widehat{i}_L\right)$ in the event that the Dictator made a promise to cooperate is that the Recipient will choose $i = 1$ and the Dictator will cooperate. Proposition 5 summarizes.

Proposition 5 *With legal enforcement the equilibrium is as follows:(i) If the Dictator made no promise to cooperate, then the Recipient will not invest and the Dictator will not cooperate,*

[44]Note that our model rules out the possibility of states of the world where cooperation is inefficient. Hence, contrary to the classic models of Shavell (1980, 1984), Goetz and Scott (1980), Cooter and Eisenberg (1985) and others, expectation damages won't produce overinvestment in our model.

$\left(\widehat{a}_L, \widehat{i}_L\right) = (0,0)$. *(ii) If the Dictator made a promise to cooperate, parties achieve the first-best,* $\left(\widehat{a}_L, \widehat{i}_L\right) = (a^*, i^*) = (1, e)$.

Proof. Because there is no legal enforcement in the absence of a promise, (i) follows immediately from Proposition 2. As for (ii), if the Dictator makes a promise, and the Recipient invests 0, there is no legal enforcement and the Dictator will choose not to cooperate. But if the Recipient chooses positive investment, the promise is enforced, and it follows from Assumption 5 that his payoff will be $\pi_R(1,i)$ no matter which action the Dictator chooses. As $\max_{i\in[0,i^c]} \pi_R(1,i) = \pi_R(1,e) > \pi_R(0,0)$ the Recipient's dominant strategy is to chose the first-best investment level e. The Dictator's payoff is therefore written as $\pi_D^L(1,e) = W^* - \pi_R(1,e)$ if she cooperates and $\pi_D^L(0,e) = W(0,e) - \pi_R(1,e)$ if she chooses not to cooperate. By the definition of W^* it must be that $\pi_D^L(1,e) > \pi_D^L(0,e)$, so that it is optimal for the Dictator to cooperate, $\left(\widehat{a}_L, \widehat{i}_L\right) = (1,e)$. ■

In other words, when the legal regime is in place, and the Dictator made a promise, the Recipient can rely on legal enforcement to give the Dictator sufficient reason to perform. He therefore does not need to overinvest in order to psychologically lock in the Dictator. Thus, legal enforcement enables the agents to achieve the first best when a promise was made.

Incomplete Information about the Dictator's Type

So far, we have been assuming that the Recipient knows the Dictator's preferences. In reality, it is likely that the Dictator is drawn from a population of agents with heterogenous preferences, and that the Recipient will therefore be uncertain about the preferences, in particular, the guilt parameter, of the particular Dictator he faces. The above analysis continues to apply to such a situation if we assume that the Recipient assigns a point estimate of γ to the Dictator so that he acts as if he knows the Dictator's type. But it is more natural to represent a Recipient's beliefs by a probability distribution over the possible Dictator types.

If the Recipient is uncertain about the Dictator's type, then, in the absence of a legal regime, the expected utility of the Recipient, if he has received a promise, is given by:

$$\begin{aligned} U_R(i) &= E[\pi_R(a,i)] \\ &= P(\gamma < \gamma^c(i))\,\pi_R(0,i) + (1 - P(\gamma < \gamma^c(i)))\,\pi_R(1,i) \\ &= \pi_R(1,i) - P(\gamma < \gamma^c(i))\,(\pi_R(1,i) - \pi_R(0,i)), \end{aligned} \tag{21}$$

where $P(\gamma < \gamma^c(i))$ is a function describing the probability he assigns to the Dictator's guilt parameter being smaller than the critical value that would induce her to cooperate. Notice that the probability a recipient should rationally assign to cooperation, $1 - P(\gamma < \gamma^c(i))$, is decreasing in $\gamma^c(i)$, and therefore rising in i. Overinvesting ($i > e$) therefore has two countervailing effects on the Recipient's expected utility. It reduces his utility for any given action of the dictator. But it might also increase the likelihood that the dictator will cooperate. This is because, increasing investment reduces $\gamma^c(i)$, the critical value of the guilt parameter beyond which the dictator will rationally cooperate. Assuming that the distribution of Dictator's sensitivity to guilt assigns positive probability mass to Dictators with sufficiently high guilt parameters, Proposition 6 immediately follows.

Proposition 6 *If the Dictator made a promise, then in the absence of the legal regime: (i) cooperation rates will increase as investment increases; and (ii) the Recipient will assign a higher probability to cooperation as investment increases.*

Example. Suppose that material payoffs are given by our particular specification in (15) and (16), and that the Dictator's guilt function is linear so that her utility in the event that she made a promise is given by:

$$u_D(a,1) = \begin{cases} 15 - 3a - \gamma(1-a)6 & \text{if } i \leq 1 \\ 15 - 3a - \gamma(1-a)\left(6.50 + .75i\right) & \text{if } i > 1 \end{cases}.$$

Suppose that the Dictator can be one of three types: a high guilt type with $\gamma = \frac{1}{2} > \frac{3}{7.25}$, a low-guilt type with $\gamma = \frac{1}{4} < \frac{3}{11}$, and an intermediate type with $\gamma = \frac{3}{8}$. The Recipient is uncertain about the Dictator's guilt parameter, assigning probabilities p_1, p_2, and p_3 to the low-, intermediate-, and high-guilt types respectively where $p_1 + p_2 + p_3 = 1$. It then follows from Proposition (4) that in the absence of a legal regime, the Dictator never cooperates if $\gamma = \frac{1}{4}$, the Dictator always cooperates regardless of the investment level if $\gamma = \frac{1}{2} > \frac{3}{7.25}$, and the Dictator cooperates if and only if $i \geq 2$ if $\gamma = \frac{3}{8}$. Thus, the likelihood that the Recipient assigns to the Dictator cooperating is p_3 if he invests less than 2, and $p_2 + p_3$ if he invests more than or equal to 2.

Since the behavior of the low-guilt and high-guilt Dictator doesn't depend on the Recipient's investment level, we can characterize the equilibrium by the Recipient's investment choice, $\widehat{i}$, and the action of the Dictator with the intermediate guilt parameter, $\widehat{a}^m$. Proposition 7 summarizes**.**

Proposition 7 *Equilibrium,* $\left(\widehat{a}^m, \widehat{i}\right)$*, in the absence of a legal regime:*
If the Dictator made no promise to cooperate, then the Recipient invests zero and the Dictator doesn't cooperate: $\left(\widehat{a}^m, \widehat{i}\right) = (0,0)$*. If the Dictator made a promise to cooperate, then the Dictator never cooperates if* $\gamma = \frac{1}{4}$*, the Dictator always cooperates irrespective of the investment level if* $\gamma = \frac{1}{2}$*, and the Dictator cooperates if* $i \geq 2$ *if* $\gamma = \frac{3}{8}$ *and does not cooperate otherwise.*
(i) If the Recipient is not sufficiently confident that the Dictator is the high-guilt type, $p_3 < 4\left(p_1 + p_2\right)$*, then he overinvests,* $\widehat{i} = 2$*, if he is sufficiently confident that the Dictator is the intermediate type,* $p_1 \leq 3p_2$*, so that* $\left(\widehat{a}^m, \widehat{i}\right) = (1,2)$*, and he invests zero,* $\widehat{i} = 0$*, so that* $\left(\widehat{a}^m, \widehat{i}\right) = (0,0)$ *otherwise.*
(ii) If the Recipient is sufficiently confident that the Dictator is of the high-guilt type, $p_3 \geq 4\left(p_1 + p_2\right)$*, then he overinvests,* $\widehat{i} = 2$*, if he is sufficiently confident that the Dictator is the intermediate type,* $p_1 \leq -\frac{1}{3} + \frac{29}{3}p_2$*, so that* $\left(\widehat{a}^m, \widehat{i}\right) = (1,2)$*, and he invests one,* $\widehat{i} = 1$*, so that* $\left(\widehat{a}^m, \widehat{i}\right) = (0,1)$ *otherwise.*

Proof. If the Dictator made no promise it follows immediately from Proposition 4 that the equilibrium is $(0,0)$. If the dictator made a promise, the Recipient's expected utility is given by:

$$U_R(i) = \begin{cases} \left(p_1 + p_2\right)(6 - i) + p_3\left(12 + 0.25i\right) & i \leq 1 \\ \left(p_1 + p_2\right)(6 - i) + p_3\left(12.50 - 0.25i\right) & 1 < i < 2 \\ p_1\left(6 - i\right) + \left(p_2 + p_3\right)\left(12.50 - 0.25i\right) & i \geq 2 \end{cases}.$$

Notice that

$$U_R(2) - \lim_{i \to 2^-} U_R(i) = 4p_1 + 12(p_2 + p_3) - 4(p_1 + p_2) - 12p_3 = 8p_2,$$

which is strictly positive so long as $p_2 > 0$. Thus, there is a discontinuity in $U_R(i)$. Notice also that

$$\frac{\partial U_R(i)}{\partial i} = \begin{cases} -(p_1 + p_2) + 0.25p_3 & i \leq 1 \\ -(p_1 + p_2) - 0.25p_3 & 1 < i < 2 \\ -p_1 - 0.25(p_2 + p_3) & i \geq 2 \end{cases}.$$

Thus, except at $i = 2$, the function is everywhere decreasing if $p_3 < 4(p_1 + p_2)$, and increasing when $i \leq 1$ and everywhere decreasing when $i > 1$ if $p_3 \geq 4(p_1 + p_2)$.
(i) First suppose that $p_3 < 4(p_1 + p_2)$. The Recipient's optimal investment is given by:

$$\widehat{i} = 2 \text{ if } U_R(2) \geq U_R(0) \Leftrightarrow p_1 \leq 3p_2$$

and

$$\widehat{i} = 0 \text{ otherwise.}$$

(ii) Now suppose that $p_3 \geq 4(p_1 + p_2)$. The Recipient's optimal investment is given by:

$$\widehat{i} = 2 \text{ if } U_R(2) \geq U_R(1) \Leftrightarrow p_1 \leq 7p_2 - 0.25p_3$$

and

$$\widehat{i} = 1 \text{ otherwise.}$$

Using the fact that $p_3 = 1 - p_1 - p_2$, we can write the Recipient's optimal investment as:

$$\widehat{i} = 2 \text{ if } U_R(2) \geq U_R(1) \Leftrightarrow p_1 \leq -\frac{1}{3} + \frac{29}{3}p_2$$

and

$$\widehat{i} = 1 \text{ otherwise.}$$

■

When the Recipient is sufficiently confident that the Dictator is an intermediate type, holding constant the likelihood that the Dictator is a high-guilt type, then he overinvests in equilibrium, investing 2, in order to induce the intermediate type to cooperate. But the Recipient's confidence that the Dictator is a high-guilt type also matters, for when he is confident that he faces a high-guilt type, and there is a sufficiently low chance that he faces an intermediate-guilt type, it is likely that the Dictator will cooperate regardless of his investment level, making it optimal for him to invest one. But if the Recipient believes instead that there is a high chance that he faces a low-guilt type, and a sufficiently low chance that he faces an intermediate-guilt type, then it is likely the Dictator won't cooperate regardless of his investment level, making it optimal for the Recipient to invest zero.

In the presence of a legal regime with Expectation Damages, all three types of Dictator cooperate so long as the Recipient invests one and the Dictator made a promise to cooperate. Thus, the uncertainty about the Dictator's type makes no difference to the equilibrium. Proposition 8 follows immediately and so is stated without proof.

Proposition 8 *Equilibrium,* $\left(\widehat{a}_L, \widehat{i}_L\right)$, *when relied-upon promises are enforced with expectation damages: (i) If the Dictator made no promise to cooperate, then the Recipient invests zero and the Dictator doesn't cooperate,* $\left(\widehat{a}_L, \widehat{i}_L\right) = (0, 0)$. *(ii) If the Dictator made a promise to cooperate, then the Recipient invests one and the Dictator cooperates,* $\left(\widehat{a}_L, \widehat{i}_L\right) = (1, 1)$.

The Communication Stage

To complete our analysis of the game, we need to examine what happens at the communication stage, which is described in detail in Step 1 of Section 3.2. Because any promise by Participant 1 is conditional on receiving a promise from Participant 2 in return, both agents have reason to make promises to one another during the communication stage so long as they anticipate that the benefits from receiving them exceed the costs of making them. The question then becomes whether there are indeed positive expected net benefits from reciprocal promises in each regime.

In No Regime, promising is cheap talk in the sense that there is no material cost associated with making a promise. But our model supposes that promisors are internally motivated to keep their promises if their guilt parameters are sufficiently high and the promisee has relied enough on their promises, such that making a promise may mean that the Dictator ends up forgoing the material gains that can be made from defecting—gains that she would have no problem realizing had she made no promise. Moreover, the specification of payoffs we used in the experiment is such that the material benefits of cooperation to the Recipient exceed the material gains the Dictator realizes from defecting, such that the expected material payoff of each agent prior to learning her role is higher if both agents cooperate in the subsequent subgame when selected to be the Dictator than if neither do. Thus, so long as each agent believes that the other agent has a sufficiently high guilt parameter to yield an equilibrium in the subsequent subgame in which the other keeps his promise if selected to be the Dictator, there are positive expected net benefits from reciprocal promises in No Regime.[45]

In Expectation Damages, the analysis is simpler. The legal regime rather than agents' internal motivations ensures that agents have sufficient reasons to keep their promises if selected to be the Dictator. The structure of the experimental payoffs then ensures that there are positive expected net benefits from reciprocal promises.[46]

[45]The magnitude of the expected net benefits from reciprocal promises will depend on the agents' guilt parameters and the resulting equilibria in the two possible subgames (the subgame where Participant 1 is Dictator and the subgame where Participant 2 is Dictator). But so long as both believes the other has a high enough guilt parameter to keep her promises in equilibrium, the expected net benefits from reciprocal promises should always be positive given the experimental payoffs we chose. This is because the cost of not being able to realize the gains of defection when the Dictator is always $\pi_D(0) - \pi_D(1) = \$15 - \$12 = \$3$, while the benefit of being a Recipient when the Dictator has made a promise is $\pi_R(1, \widehat{i}) - \pi_R(0, 0) = \$6.25 - 0.25\left|1 - \widehat{i}\right| \geq \$5 > \$3$, given that the maximum possible investment level is 6.

[46]A formal model of the communication stage is available upon request.

ONLINE APPENDIX B: Alternative Hypotheses

Here we consider whether competing behavioral theories like inequity aversion or efficiency preferences could be driving our results instead of our model of promisors' preferences. We explain how the comparison between No Regime and Control support our claim that our posited promissory mechanism lies behind our results.

Efficiency Preferences and Equity Preferences

Efficiency Preferences. By equations (15) and (16), the difference in the agents' joint payoffs when the Dictator cooperates, $a = 1$, and when she doesn't cooperate, $a = 0$, is $3 + \frac{5}{4}$ when $i \leq 1$ and $3.5 + \frac{3}{4}i$ when $i > 1$. Thus, the difference in social surplus is higher when the Dictator cooperates than when she doesn't, and the difference is increasing in the Recipient's investment i. If the Dictator has a preference for efficiency (i.e., she believes that a higher social surplus is better) that she trades off against her material payoff, then she will be more likely to cooperate the higher is the Recipient's investment i. Hence, we could see the results predicted by Hypothesis 2 even if there was no promise-induced cooperation.

Equity Preferences. By equations (15) and (16), the payoff differences between the Dictator and the Recipient when the Dictator cooperates, $a = 1$, and when she doesn't cooperate, $a = 0$, and the difference between the two differences, Δ, are as follows:

	$a = 1$	$a = 0$	Δ
$i \leq 1$	$-\frac{1}{4}i$	$9 + i$	$9 + \frac{5}{4}i$
$i > 1$	$-0.5 + \frac{1}{4}i$	$9 + i$	$9.5 + \frac{3}{4}i$

Thus, the difference between the payoffs is smaller when the Dictator cooperates than when she doesn't, and the difference between the differences is increasing in the Recipient's investment i. If the Dictator has a preference for equity that she trades off against her material payoff, then she will be more likely to cooperate the higher is the Recipient's investment i. Hence, again, we could see the results predicted by Hypothesis 3 even if there was no promise-induced cooperation.

These competing hypotheses do not challenge our basic claim that reliance investments are used strategically by recipients to make their dictators' cooperation more likely. Nor do they challenge our central insight that legal enforcement of promises can reduce overinvestment as well as underinvestment. But they raise doubts as to whether promising is a reason why recipients' reliance investments make dictators more willing to keep their promises.

Results from the Control Treatment

The results of an additional Control treatment, which is identical to the No Regime treatment with the exception that participants are not able to send messages to each other, provide direct empirical support for the claim that there is something distinctively *promissory* about our lock-in effect. By this we mean that a dictator's sensitivity to her recipient's investment choice is triggered, or, at the very least intensified, when she made a promise.

To be more precise, let's generalize our model to allow for the possibility that equity or efficiency concerns could be creating a lock-in effect. That is, let's suppose that the Dictator's utility function is given by:

$$u_D(a, p) = \pi_D(a) - p\gamma g\left(\pi_R(1, i) - \pi_R(a, i)\right) + \sigma q(a, i),$$

where $q(1,i) - q(0,i)$ is an increasing function in i, and $\sigma \geq 0$ is a parameter that reflects the magnitude of the Dictator's equity and efficiency concerns, such that if $\sigma > 0$, then we would have an alternative explanation for our psychological lock-in effect.[47] Thus, the difference in the Dictator's utility when she cooperates compared to her utility when she does not cooperate is given by:

$$\pi_D(1) - \pi_D(0) + p\gamma g(\pi_R(1,i) - \pi_R(0,i)) + \sigma(q(1,i) - q(0,i)),$$

which is also increasing in i. It follows that, if there is a promissory effect of investment on the Dictator's willingness to cooperate, i.e. if $\gamma > 0$, then the critical value of investment beyond which she is willing to cooperate, $i^c(p)$, will be lower if the Dictator made a promise: $i^c(1) < i^c(0)$.

Now suppose, in the spirit of the imperfect information version of our basic model, that there is a continuum of possible dictators, each of whom can be characterized by such a pair of critical values so that the jth dictator can be characterized by the vector $\left(i_j^c(1), i_j^c(0)\right)$. If the Recipient doesn't know the preferences of the Dictator he faces, all he sees are two probability distributions of critical values, $F_1(x)$ and $F_0(x)$ where $F_1(x) = \Pr\left[i_j^c(1) \leq x\right]$ and $F_0(x) = \Pr\left[i_j^c(0) \leq x\right]$ and $f_1(x)$ and $f_0(x)$ are the corresponding probability density functions. Thus, if there is a promissory lock-in effect, i.e. $\gamma > 0$, then it must always be the case that $F_1(x) \geq F_0(x)$ for all x, and, so long as $F_0(0) < 1$, then $F_1(x) > F_0(x)$ for some $x \in [0,6]$. In other words, given a promissory lock-in effect, a promise will shift the distribution function in the sense of first-order stochastic dominance.

If there is an equity or efficiency lock-in effect, i.e. $\sigma > 0$, then, even in the absence of a promise, the Recipient's confidence that the Dictator will cooperate will be non-decreasing in investment, and it should be strictly increasing over any range of investment levels where there is probability mass, $f_0(x) > 0$. Thus, one piece of evidence against the equity/efficiency concerns explanation of our lock-in effect is that the average confidence of recipients that a dictator will cooperate is not increasing in the investment level in Control. There is no statistically significant difference between the average recipient confidence levels at investment levels zero and six (p=0.89, Wilcoxon sign-rank test; p=0.39, one-sided t-test). Indeed, the graph shows a statistically significant decrease in recipient confidence from $i = 0$ to $i = 1$ in Control (p=0.02, Wilcoxon sign-rank test). By contrast, in No Regime, average recipient confidence is significantly higher for investment level six than investment level zero (Wilcoxon

[47] By supposing that the three component's of the Dictator's utility function are additive, we assume that the Dictator's promissory concerns are separable from her efficiency or equity concerns. If preferences have this structure, then if $\sigma > 0$, efficiency or equity concerns would generate psychological lock-in in the absence of a promise ($\gamma = 0$). Likewise, if $\gamma > 0$, promising would generate psychological lock-in in the absence of efficiency or equity concerns ($\sigma = 0$).

An alternative possibility is that the effects of promising and equity/efficiency concerns are not independent of one another such that *both* have to be present for psychological lock-in to occur. For instance, the Dictator's utility might instead take the form: $u_D(a,p) = \pi_D(a) + \sigma p\gamma f(a,i)$. In that case, in the absence of a promise, or in the absence of equity concerns, there would be no possibility of psychological lock-in.

Our comparison of Control and No Regime allows us to investigate whether there is a promissory lock-in effect. But we can't use the comparison to determine whether the promissory effect arises independently of an equity/efficiency based lock-in effect or instead acts only in conjunction with an equity/efficiency effect. To distinguish these two possible explanations of our data, we would have to construct an experiment with payoffs that ensured that potential equity/efficiency concerns were not increasing in the investment level.

sign-rank test, p=0.04), and the graph of recipient confidence is everywhere non-decreasing (Figure OA1).

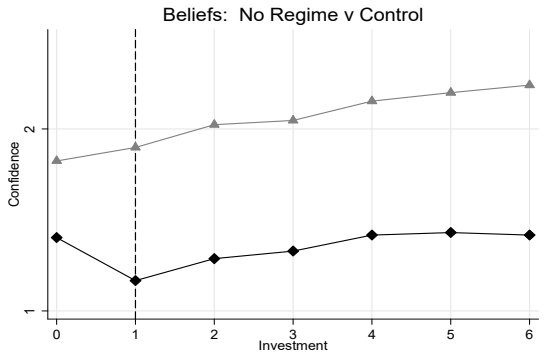


Figure OA1

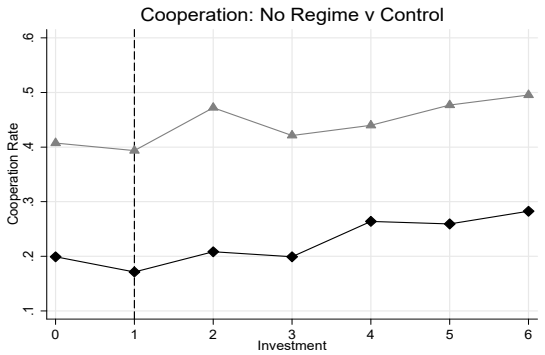


Figure OA2

Consistent with this belief data, moreover, more recipients overinvest in No Regime compared to Control: 22.7% compared to 13.9%. This effect is statistically significant (Wilcoxon rank sum, p=0.08, one-sided t-test p=0.03). Thus, we have evidence that recipients believe there is a lock-in effect under No Regime but not under Control, and recipients' beliefs are ultimately what drives their investment behavior. Moreover, particularly given that our recipients sometimes find themselves in the role of Dictator (given that the experiment randomly allocates the roles across the eight rounds of the experiment), this belief data ought to be informative about the underlying preferences of dictators.

Our contingent cooperation data (Figure OA2), however, don't paint as clear a picture. On the one hand, cooperation rates strongly increase for No Regime between the efficient investment level, $i = 1$, and the lowest level of overinvestment, $i = 2$, and the difference is significant (p=0.02, Wilcoxon sign-rank test), while there is barely a significant difference between cooperation rates under Control over this range (p=0.20, Wilcoxon sign-rank; p=0.10, one-sided t-test). On the other hand, the difference in cooperation rates at the investment boundaries is significant under Control (p=0.02, Wilcoxon sign rank test), suggesting that there is some equity/efficiency driven lock-in effect, while the difference is only significant at 10% in No Regime on the one-sided t-test (p=0.18 Wilcoxon sign-rank test; p=0.07, one-sided t-test). But contingent cooperation data may not be that reliable at higher levels of investment given that few people choose the higher investment levels so dictators don't have to worry so much about the consequences of those cooperation decisions. In effect, dictators' decisions at lower levels of investment are much less hypothetical than those at higher levels. Roughly 92% of the Recipients invest two or less under Control and roughly 88% invest two or less under No Regime. Under both treatments roughly 93% invest three or less.

It might seem odd that the belief and contingent cooperation data don't track one another more closely given that all subjects experience both roles throughout the experiment. But this could arise from the fact that dictators' cooperation decisions are binary at each investment level while recipients' statements of beliefs could take on one of five possible values, which could have introduced more noise into the cooperation data. Also, as we already noted, dictators may not take their cooperation decisions seriously at higher investment levels given the low chance that such decisions have payoff consequences, which would also reduce the reliability of the cooperation data at the higher investment levels. By contrast, recipients

were paid for their predictions about dictator behavior at higher levels of investment just as often as they were paid for their predictions about dictator behavior at lower levels.

ONLINE APPENDIX C: POST EXPERIMENT QUESTIONNAIRE

1. **The question asking for subjects' understanding of the payoff structure was:**

   (a) Did you understand how A's and B's actions affected their profits? (Yes: 1079 ; No: 4; Kind of: 122).[48]

2. **The questions to assess carefulness and honesty of subjects were:**

   (a) I didn't take the scenario seriously. I just wanted to earn the $1.50 fee as quickly as possible. (Yes: 11 out of 1205);

   (b) I carefully read the instructions. (No: 6 out of 1205);

   (c) I chose my answers to make myself seem like a good person. (Yes: 171 out of 1205)

   (d) This is the first time I have completed this survey. (No: 4 out of 1205).

3. **The demographic questions were:**

   (a) What is your age? (Age was between 18 and 77 with and average age of 35.7)

   (b) What is you gender? (53% female)

   (c) What is your highest level of schooling? (Master's degree or more: 12.2%; Bachelor's degree: 39.3%; Associate's degree: 15.7%; Vocational or technical certificate/ diploma after high school (such as cosmetics): 8.4%; Highschool diploma: 24%; I did not complete Highschool: 0.5%)

   (d) Is English your first language? (No: 19 out of 1205).

[48]Note that after reading the instructions but before subjects were allowed to start the study they had to answer control questions about the the payoffs in the vignette and were only allowed to proceed once they had answered them correctly. Nonetheless, some subjects reported that they were confused or unsure. We introduced the "Kind of" category to be conservative. We assumed that many subjects would not be willing to admit that they did not understand the payoffs at all.

ONLINE APPENDIX D: INSTRUCTIONS*[49]

[49]The first 5 pages are the instructions for the No Regime treament and second 5 pages are the instructions for the Expectation Damages treatment.

Instructions

Thank you for participating in this experiment. The purpose of this experiment is to study how people make decisions in certain situations. If you have questions at any time, please raise your hand. Please do not speak to other participants during the experiment. Depending on the decisions made by you and other participants, you may earn money as described below. The money that you make from this experiment will be added to the $5 show-up fee and the money you will earn from completing a post-experiment survey and the total will be paid to you individually and privately at the end of the experiment.

This session consists of 2 practice rounds and 8 paying rounds in which you may earn money. In each round, you will interact with another randomly chosen participant. **Under no circumstances will you interact with the same participant twice in a paying round**; that is, in every paying round you will be paired with a participant that you have not been paired with before. No participant will learn the identity of the persons with whom he or she interacted during any of the rounds.

At the end of the experiment, **one of the 8 paying rounds in each session will be randomly chosen for payment (every round is equally likely to be chosen)**. The amount that you will receive at the end of the experiment will depend on the decisions made in that round.

Overview. You will be randomly assigned one of two roles: Player A or Player B. At the end of the game, Player A gets to decide whether or not to send money to Player B. Before roles are assigned, both participants can decide whether or not to make a promise to send money to Player B if they are assigned the role of Player A. Once roles are assigned, Player B makes an investment decision that may affect his own payoff. Finally, Player A decides whether or not to send money to Player B.

Each round consists of four steps, which are described below.

Step 1: Communication Phase. At the beginning of each round, you will be anonymously and randomly matched with an interaction partner. You will then have the opportunity to exchange messages with one another sequentially. The order in which you make these communication decisions will be determined randomly. The participant who is selected to make the first communication decision (Participant 1) must decide whether or not to send Message 1 to the other participant (Participant 2):

Message 1: “I promise to send you money if I am chosen to be Player A so long as you make me a return promise.”

After learning of Participant 1’s decision, Participant 2 must decide whether or not to send Message 2 to Participant 1:

Message 2: “I promise to send you money if I am chosen to be Player A.”

Table 1 describes the consequences of the decisions that the participants may make during this communication phase. Notice that Participant 1 only ends up promising to send Participant 2 money if Participant 2 makes a promise in return.

Table 1: The consequences of the Communication Phase

	Participant 1 sends Message 1	Participant 1 doesn't send Message 1
Participant 2 sends Message 2	**Both** participants have promised	**Only** Participant 2 has promised
Participant 2 doesn't send Message 2	**Neither** participant has promised	**Neither** participant has promised

Step 2: Role Assignment. Once the Communication Phase is over, you and your interaction partner will learn whether you have been chosen to be Player A or Player B (you will learn more about the meaning of these roles below). Your role will be randomly assigned anew in each round. It is always equally likely that you will be Player A or Player B (regardless of the messages you send or the actions you take in any of the rounds).

Step 3: Player B's Investment Decision. Once roles have been assigned, Player B must make an investment decision, which may influence his own payoff. Exactly how the investment decision affects his payoff depends on the action Player A chooses at the next step (you will learn more about how payoffs are determined below).

Step 3a: Player B's Guessing. Player B has the opportunity to earn bonus money by indicating how likely he thinks it is that Player A will decide to send him money at the next stage of the experiment for each level of investment that he might choose. Thus, prior to making his investment decision, Player B should indicate the likelihood with which he believes that Player A will send him money if he invests 0, if he invests 1, if he invests 2, and so on.

More specifically, for each possible investment level that Player B may choose, Player B must indicate which of the following best approximates his level of confidence that Player A will send him money: Player A will certainly send him money; Player A will probably send him money; there is a 50-50 chance that Player A will send him money; Player A probably won't send him money; Player A certainly won't send him money.

One round that is *not* chosen for payment based on the participants' actual decisions will be randomly selected for payment based on Player B's guesses in that round. Thus, if you were Player B in the randomly selected round, you will be paid for your guesses in that round.

Table 2 shows how Player B's payoff from guessing is determined. For example, if Player B states that he believes that Player A will certainly send him money for a particular investment level, and Player A decides to send him money if he chooses that investment level, then Player B gets a payoff of $0.65. But if Player A decides not send Player B money for that investment level, then Player B gets a payoff of $0.15.

Table 2: How Player B's earnings depend on B's stated beliefs about Player A's action.

	Player A will certainly send Player B money	Player A will probably send Player B money	There is a 50-50 chance that Player A sends Player B money	Player A probably will not send Player B money	Player A certainly will not send Player B money
Player B's earnings if Player A decides to send him/her money	$0.65	$0.60	$0.50	$0.35	$0.15
Player B's earnings if Player A decides not send Player B money	$0.15	$0.35	$0.50	$0.60	$0.65

Step 3b: Player B's Decision. After Player B has indicated his level of confidence that Player A will send him money for each possible investment level, Player B chooses his investment level.

Step 4: Player A's Decision. Once Player B has made his investment decision, Player A must decide whether or not to send Player B money. Player A must make this decision before she learns what investment decision Player B actually made.

Player A makes her decision by indicating what she would like to do for each level of investment that Player B could have chosen. That is, Player A must indicate whether or not she will send Player B money if his investment was 0, whether or not she will send him money if his investment was 1, and so on.

Player A's actual choice (and therefore the final payoffs of Player A and Player B) will depend on the investment level that Player B has chosen. So, for example, if Player B chose an investment level of 3, Player A sends him money if she indicated that she would do so for that investment level.

Payoffs. Player A's decision whether or not to send Player B money in conjunction with Player B's investment decision determine both participants' payoffs.

Tables 3 and 4 illustrate how the players' payoffs depend on these decisions.

Table 3: Player A **sends** Player B money	B's Investment	A's Payoff	B's Payoff
	0	$12	$12.00
	1	$12	$12.25
	2	$12	$12.00
	3	$12	$11.75
	4	$12	$11.50
	5	$12	$11.25
	6	$12	$11.00

Table 4: Player A **doesn't send** Player B money	B's Investment	A's Payoff	B's Payoff
	0	$15	$6
	1	$15	$5
	2	$15	$4
	3	$15	$3
	4	$15	$2
	5	$15	$1
	6	$15	$0

Preliminary Questions. Before the experiment begins, you will be asked two questions to check that you understand how Player A's and Player B's decisions determine their payoffs.

Final Questions. At the end of the experiment, you will be asked to answer some further questions, and you will have the opportunity to make some additional money.

Do you have any questions?

Preliminary Questions

1) What are the payoffs of Player A and Player B if Player A chooses not to send Player B money and Player B invests 4?

Player A's payoff	
Player B's payoff	

2) What are the payoffs of Player A and Player B if Player A chooses to send Player B money and Player B invests 1?

Player A's payoff	
Player B's payoff	

Instructions

Thank you for participating in this experiment. The purpose of this experiment is to study how people make decisions in certain situations. If you have questions at any time, please raise your hand. Please do not speak to other participants during the experiment. Depending on the decisions made by you and other participants, you may earn money as described below. The money that you make from this experiment will be added to the $5 show-up fee and the money you will earn from completing a post-experiment survey and the total will be paid to you individually and privately at the end of the experiment.

This session consists of 2 practice rounds and 8 paying rounds in which you may earn money. In each round, you will interact with another randomly chosen participant. **Under no circumstances will you interact with the same participant twice in a paying round**; that is, in every paying round you will be paired with a participant that you have not been paired with before. No participant will learn the identity of the persons with whom he or she interacted during any of the rounds.

At the end of the experiment, **one of the 8 paying rounds in each session will be randomly chosen for payment (every round is equally likely to be chosen)**. The amount that you will receive at the end of the experiment will depend on the decisions made in that round.

Overview. You will be randomly assigned one of two roles: Player A or Player B. At the end of the game, Player A gets to decide whether or not to send money to Player B. Before roles are assigned, both participants can decide whether or not to make a promise to send money to Player B if they are assigned the role of Player A. Subsequently, Player B can make an investment that may affect the payoffs of both participants. Finally, Player A decides whether or not to send money to Player B.

Each round consists of four steps, which are described below.

Step 1: Communication Phase. At the beginning of each round, you will be anonymously and randomly matched with an interaction partner. You will then have the opportunity to exchange messages with one another sequentially. The order in which you make these communication decisions will be determined randomly. The participant who is selected to make the first communication decision (Participant 1) must decide whether or not to send Message 1 to the other participant (Participant 2):

Message 1: “I promise to send you money if I am chosen to be Player A so long as you make me a return promise.”

After learning of Participant 1’s decision, Participant 2 must decide whether or not to send Message 2 to Participant 1:

Message 2: “I promise to send you money if I am chosen to be Player A.”

Table 1 describes the consequences of the decisions that the participants may make during this communication phase. Notice that Participant 1 only ends up promising to send Participant 2 money if Participant 2 makes a promise in return.

Table 1: The consequences of the Communication Phase

	Participant 1 sends Message 1	Participant 1 doesn't send Message 1
Participant 2 sends Message 2	**Both** participants have promised	**Only** Participant 2 has promised
Participant 2 doesn't send Message 2	**Neither** participant has promised	**Neither** participant has promised

Step 2: Role Assignment. Once the Communication Phase is over, you and your interaction partner will learn whether you have been chosen to be Player A or Player B (you will learn more about the meaning of these roles below). Your role will be randomly assigned anew in each round. It is always equally likely that you will be Player A or Player B (regardless of the messages you send or the actions you take in any of the rounds).

Step 3: Player B's Investment Decision. Once roles have been assigned, Player B must make an investment decision, which may influence both players' payoff. Exactly how the investment decision affects the players' payoffs depends on the action Player A chooses at the next step (you will learn more about how payoffs are determined below).

Step 3a: Player B's Guessing. Player B has the opportunity to earn bonus money by indicating how likely he thinks it is that Player A will decide to send him money at the next stage of the experiment for each level of investment that he might choose. Thus, prior to making his investment decision, Player B should indicate the likelihood with which he believes that Player A will send him money if he invests 0, if he invests 1, if he invests 2, and so on.

More specifically, for each possible investment level that Player B may choose, Player B must indicate which of the following best approximates his level of confidence that Player A will send him money: Player A will certainly send him money; Player A will probably send him money; there is a 50-50 chance that Player A will send him money; Player A probably won't send him money; Player A certainly won't send him money.

One round that is *not* chosen for payment based on the participants' actual decisions will be randomly selected for payment based on Player B's guesses in that round. Thus, if you were Player B in the randomly selected round, you will be paid for your guesses in that round.

Table 2 shows how Player B's payoff from guessing is determined. For example, if Player B states that he believes that Player A will certainly send him money for a particular investment level, and Player A decides to send him money if he chooses that investment level, then Player B gets a payoff of $0.65. But if Player A decides not send Player B money for that investment level, then Player B gets a payoff of $0.15.

Table 2: How Player B's earnings depend on B's stated beliefs about Player A's action.

	Player A will certainly send Player B money	Player A will probably send Player B money	There is a 50-50 chance that Player A sends Player B money	Player A probably will not send Player B money	Player A certainly will not send Player B money
Player B's earnings if Player A decides to send him/her money	$0.65	$0.60	$0.50	$0.35	$0.15
Player B's earnings if Player A decides not to send Player B money	$0.15	$0.35	$0.50	$0.60	$0.65

Step 3b: Player B's Decision. After Player B has indicated his level of confidence that Player A will send him money for each possible investment level, Player B chooses his investment level.

Step 4: Player A's Decision. Once Player B has made his investment decision, Player A must decide whether or not to send Player B money. Player A must make this decision before she learns what investment decision Player B actually made.

Player A makes her decision by indicating what she would like to do for each level of investment that Player B could have chosen. That is, Player A must indicate whether or not she will send Player B money if his investment was 0, whether or not she will send him money if his investment was 1, and so on.

Player A's actual choice (and therefore the final payoffs of Player A and Player B) will depend on the investment level that Player B has chosen. So, for example, if Player B chose an investment level of 3, Player A sends him money if she indicated that she would do so for that investment level.

Legal regime. There is a legal regime in place that protects Player B if Player A made a promise and Player B made an investment of at least 1. This legal regime forces Player A to compensate Player B if she breaks a promise to send Player B money (unless Player B made a zero investment decision, in which case Player A is not forced to compensate Player B). That is, if Player A made a promise to send money to Player B, and Player B invested at least 1, then the legal regime forces Player A to pay some money to Player B if Player A breaks her promise. Under this particular legal regime, Player A must pay Player B compensation that gives Player B the payoff he would have received had Player A kept her promise.

Payoffs. Player A's decision whether or not to send Player B money in conjunction with Player B's investment decision determine both participants' payoffs. Because the legal regime only forces Player A to compensate Player B if Player A breaks a promise to send Player B money, payoffs differ depending on whether or not Player A made such a promise. Tables 3 through 6 illustrate how the Players' payoffs depend on their decisions.

Payoffs if Player A **did not promise** to send Player B money

Table 3: Player A **sends** Player B money	B's Investment	A's Payoff	B's Payoff
	0	$12	$12.00
	1	$12	$12.25
	2	$12	$12.00
	3	$12	$11.75
	4	$12	$11.50
	5	$12	$11.25
	6	$12	$11.00

Table 4: Player A **doesn't send** Player B money	B's Investment	A's Payoff	B's Payoff
	0	$15	$6
	1	$15	$5
	2	$15	$4
	3	$15	$3
	4	$15	$2
	5	$15	$1
	6	$15	$0

Payoffs if Player A **promised** to send Player B money

Table 5: Player A **sends** Player B money	B's Investment	A's Payoff	B's Payoff
	0	$12	$12.00
	1	$12	$12.25
	2	$12	$12.00
	3	$12	$11.75
	4	$12	$11.50
	5	$12	$11.25
	6	$12	$11.00

Table 6: Player A **doesn't send** Player B money	B's Investment	A's Payoff	B's Payoff
	0	$15.00	$6.00
	1	$7.75	$12.25
	2	$7.00	$12.00
	3	$6.25	$11.75
	4	$5.50	$11.50
	5	$4.75	$11.25
	6	$4.00	$11.00

Preliminary Questions. Before the experiment begins, you will be asked two questions to check that you understand how Player A's and Player B's decisions determine their payoffs.

Final Questions. At the end of the experiment, you will be asked to answer some further questions, and you will have the opportunity to make some additional money.

Do you have any questions?

Preliminary Questions

1) Player A sent Message 1. Player B sent Message 2 and invested 4. What is the payoff of Player A and Player B if Player A chooses not to send Player B money?

Player A's payoff	
Player B's payoff	

2) Player A sent Message 1. Player B did not send Message 2 and invested 1. What is the payoff of Player A and Player B if Player A chooses not to send Player B money?

Player A's payoff	
Player B's payoff	

ONLINE APPENDIX E: Selected Screenshots

Period

1 out of 1

Remaining Time [sec]: 23

Now the main experiment will begin. You will be randomly matched with a different participant in this round and each successive round of the main experiment.

Continue

Period

1 out of 1

Remaining Time [sec]: 40

Do you want to send the following message to the other player?

"I promise to send you money if I am chosen to be Player A so long as you make me a return promise."

"Yes, I promise."

"No, I do NOT promise."

Period

1 out of 1

Remaining Time [sec]: 58

The other player did not make you a promise.

Do you want to send the following message to the other player?

"I promise to send you money if I am chosen to be Player A."

"Yes, I promise."

"No, I do NOT promise."

Period

1 out of 1

Remaining Time [sec]: 56

The other player has promised to send you money if he/she is chosen to be Player A so long as you make a return promise.

Do you want to send the following message to the other player?

"I promise to send you money if I am chosen to be Player A."

"Yes, I promise."

"No, I do NOT promise."

Period

1 out of 1

Remaining Time [sec]: 12

You were selected to be **Player A.**

You made **no promise** to Player B.

Player B made **no promise** to you.

Continue

Period

1 out of 2

Remaining Time [sec]: 28

You were selected to be **Player A.**

You made **no promise** to Player B.

Player B made a **promise** to you.

Continue

Period

1 out of 1

Remaining Time [sec]: 59

You were selected to be **Player A.**

You made a **promise** to Player B.

Player B made **no promise** to you.

Continue

Period

1 out of 1

Remaining Time [sec]: 18

You were selected to be **Player A.**

You made a **promise** to Player B.

Player B made a **promise** to you.

Continue

Period

1 out of 1

Remaining Time [sec]: 60

You were selected to be **Player B.**

You made **no promise** to Player A.

Player A made **no promise** to you.

Continue

Period

1 out of 1

Remaining Time [sec]: 31

You were selected to be **Player B.**

You made **no promise** to Player A.

Player A made a **promise** to you.

Continue

Period

1 out of 2

Remaining Time [sec]: 59

You were selected to be **Player B.**

You made a **promise** to Player A.

Player A made **no promise** to you.

Continue

Period

1 out of 1

Remaining Time [sec]: 59

You were selected to be **Player B.**

You made a **promise** to Player A.

Player A made a **promise** to you.

Continue

Period

1 out of 1

Remaining Time [sec]: 117

You must decide whether or not to send Player B money for each of the possible investment levels that Player B might choose.

Your actual choice and therefore final payoffs will depend on the investment level that Player B chooses. So, for example, if Player B chooses an investment level of 3, your choice is whatever you decided to do for that investment level.

You will now be asked what you want to do for each investment level that Player B might choose.

Continue

Period

1 out of 1

Remaining Time [sec]: 85

You must now make an investment decision.

Before you make your decision, please indicate your level of confidence that Player A will send you money for each possible investment level you may choose.

You will be paid depending on what Player A actually does in accordance with the table below.

	Player A will certainly send you money	Player A will probably send you money	Unsure what Player A will do	Player A probably won't send you money	Player A certainly won't send you money
Your earnings if Player A sends you money	$0.65	$0.60	$0.50	$0.35	$0.15
Your earnings if Player A doesn't send you money	$0.15	$0.35	$0.50	$0.60	$0.65

Continue

Period

1 out of 1

Remaining Time [sec]: 0

You made **no promise** to send Player B money.

Player B made **no promise** to send you money.

Suppose Player B invests 0.

Would you like to send money to Player B if Player B invests 0?

Send Money

Payoff to you	12.00
Payoff to Player B	12.00

Don't Send Money

Payoff to you	15.00
Payoff to Player B	6.00

Period

1 out of 1

Remaining Time [sec]: 0

You made **no promise** to send Player B money.

Player B made **no promise** to send you money.

Suppose Player B invests 1.

Would you like to send money to Player B if Player B invests 1?

Send Money

Payoff to you	12.00
Payoff to Player B	12.25

Don't Send Money

Payoff to you	15.00
Payoff to Player B	5.00

Period

1 out of 1

Remaining Time [sec]: 0

You made **no promise** to send Player B money.

Player B made **no promise** to send you money.

Suppose Player B invests 2.

Would you like to send money to Player B if Player B invests 2?

Send Money

Payoff to you	12.00
Payoff to Player B	12.00

Don't Send Money

Payoff to you	15.00
Payoff to Player B	4.00

Period

1 out of 1

Remaining Time [sec]: 0

Please reach a decision!

You made **no promise** to send Player B money.

Player B made **no promise** to send you money.

Suppose Player B invests 3.

Would you like to send money to Player B if Player B invests 3?

Send Money

Payoff to you	12.00
Payoff to Player B	11.75

Don't Send Money

Payoff to you	15.00
Payoff to Player B	3.00

Period

1 out of 1

Remaining Time [sec]: 0

You made **no promise** to send Player B money.

Player B made **no promise** to send you money.

Suppose Player B invests 4.

Would you like to send money to Player B if Player B invests 4?

Send Money

Payoff to you	12.00
Payoff to Player B	11.50

Don't Send Money

Payoff to you	15.00
Payoff to Player B	2.00

Period

1 out of 1

Remaining Time [sec]: 0

Please reach a decision!

You made **no promise** to send Player B money.

Player B made **no promise** to send you money.

Suppose Player B invests 5.

Would you like to send money to Player B if Player B invests 5?

Send Money

Payoff to you	12.00
Payoff to Player B	11.25

Don't Send Money

Payoff to you	15.00
Payoff to Player B	1.00

Period

1 out of 1

Remaining Time [sec]: 0

You made **no promise** to send Player B money.

Player B made **no promise** to send you money.

Suppose Player B invests 6.

Would you like to send money to Player B if Player B invests 6?

Send Money

Payoff to you	12.00
Payoff to Player B	11.00

Don't Send Money

Payoff to you	15.00
Payoff to Player B	0.00

Period

1 out of 1

Remaining Time [sec]: 55

You made a **promise** to send Player B money.

Player B made a **promise** to send you money.

Suppose Player B invests 0.

Would you like to send money to Player B if Player B invests 0?

Send Money

Payoff to you	12.00
Payoff to Player B	12.00

Don't Send Money

Payoff to you	15.00
Payoff to Player B	6.00

Period

1 out of 1

Remaining Time [sec]: 28

You made a **promise** to send Player B money.

Player B made a **promise** to send you money.

Suppose Player B invests 1.

Would you like to send money to Player B if Player B invests 1?

Send Money

Payoff to you	12.00
Payoff to Player B	12.25

Don't Send Money

Payoff to you	7.75
Payoff to Player B	12.25

Period

1 out of 1

Remaining Time [sec]: 11

You made a **promise** to send Player B money.

Player B made a **promise** to send you money.

Suppose Player B invests 2.

Would you like to send money to Player B if Player B invests 2?

Send Money

Payoff to you	12.00
Payoff to Player B	12.00

Don't Send Money

Payoff to you	7.00
Payoff to Player B	12.00

Period

1 out of 1

Remaining Time [sec]: 2

You made a **promise** to send Player B money.

Player B made a **promise** to send you money.

Suppose Player B invests 3.

Would you like to send money to Player B if Player B invests 3?

Send Money

Payoff to you	12.00
Payoff to Player B	11.75

Don't Send Money

Payoff to you	6.25
Payoff to Player B	11.75

Period

1 out of 1

Remaining Time [sec]: 0

You made a **promise** to send Player B money.

Player B made a **promise** to send you money.

Suppose Player B invests 4.

Would you like to send money to Player B if Player B invests 4?

Send Money

Payoff to you	12.00
Payoff to Player B	11.50

Don't Send Money

Payoff to you	5.50
Payoff to Player B	11.50

Period

1 out of 1

Remaining Time [sec]: 0

Please reach a decision!

You made a **promise** to send Player B money.

Player B made a **promise** to send you money.

Suppose Player B invests 5.

Would you like to send money to Player B if Player B invests 5?

Send Money

Payoff to you	12.00
Payoff to Player B	11.25

Don't Send Money

Payoff to you	4.75
Payoff to Player B	11.25

Period

1 out of 1

Remaining Time [sec]: 0

You made a **promise** to send Player B money.

Player B made a **promise** to send you money.

Suppose Player B invests 6.

Would you like to send money to Player B if Player B invests 6?

Send Money

Payoff to you	12.00
Payoff to Player B	11.00

Don't Send Money

Payoff to you	4.00
Payoff to Player B	11.00

Period

1 out of 1

Remaining Time [sec]: 88

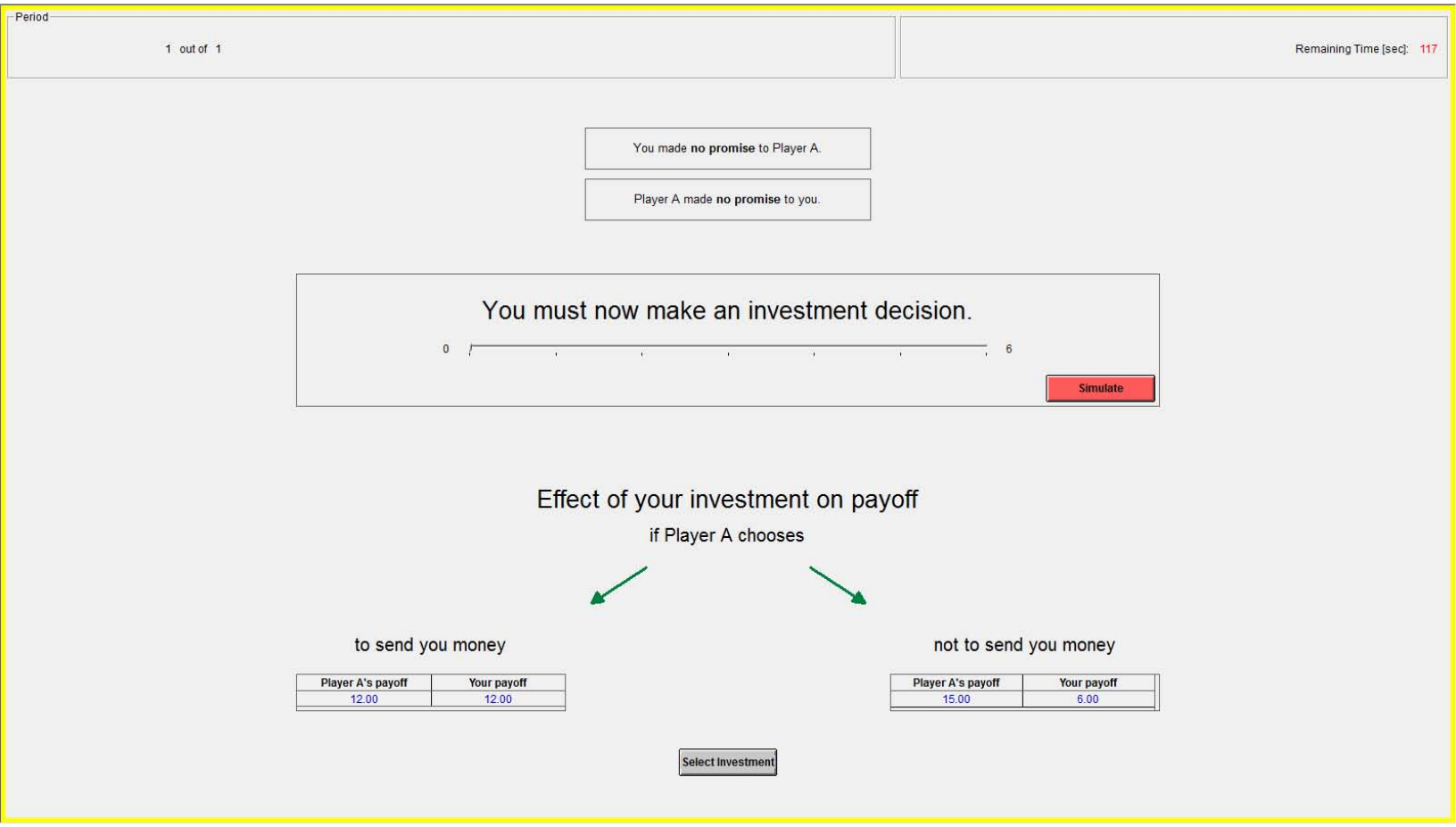
You made a **promise** to Player A.

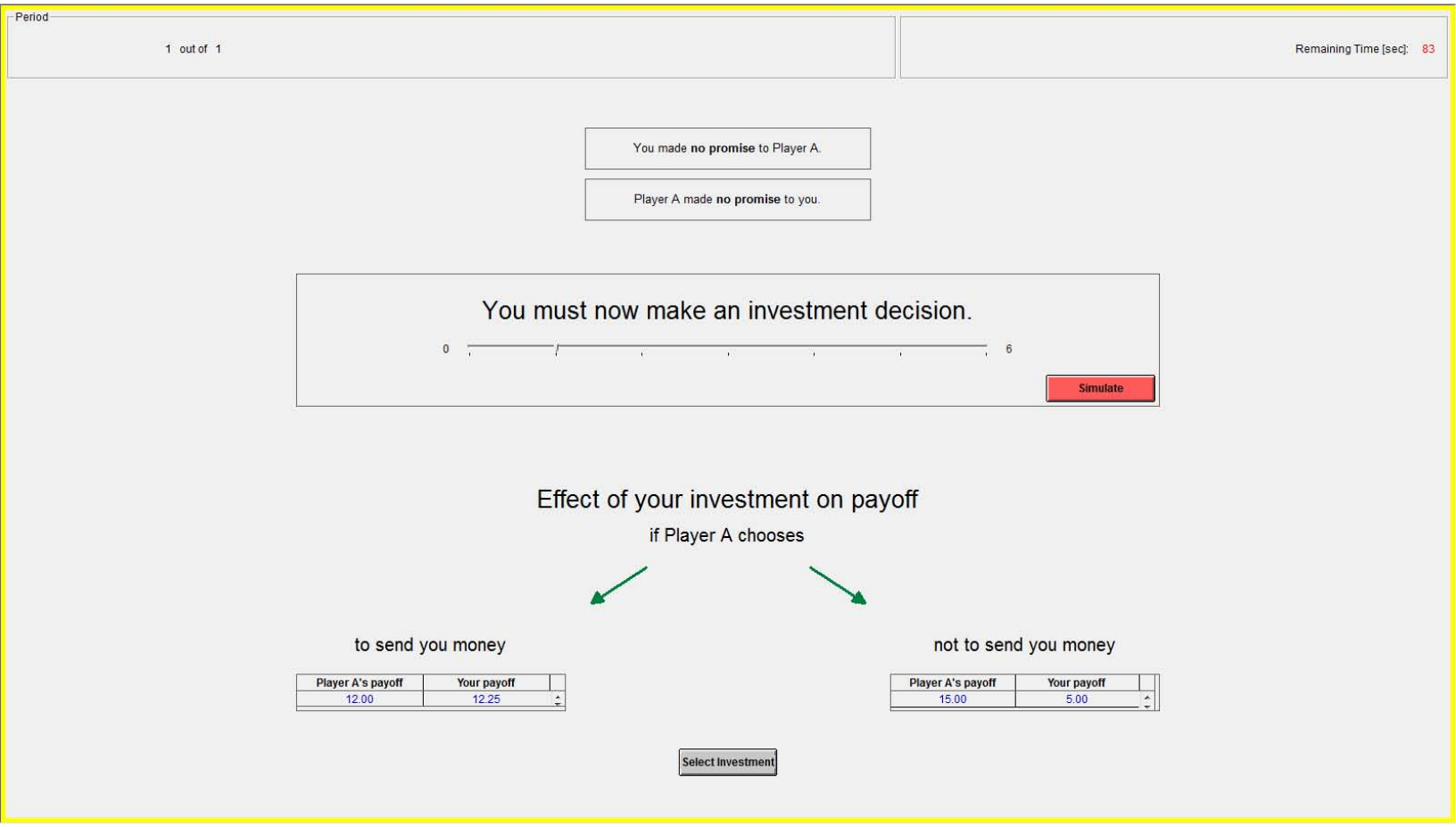
Player A made a **promise** to you.

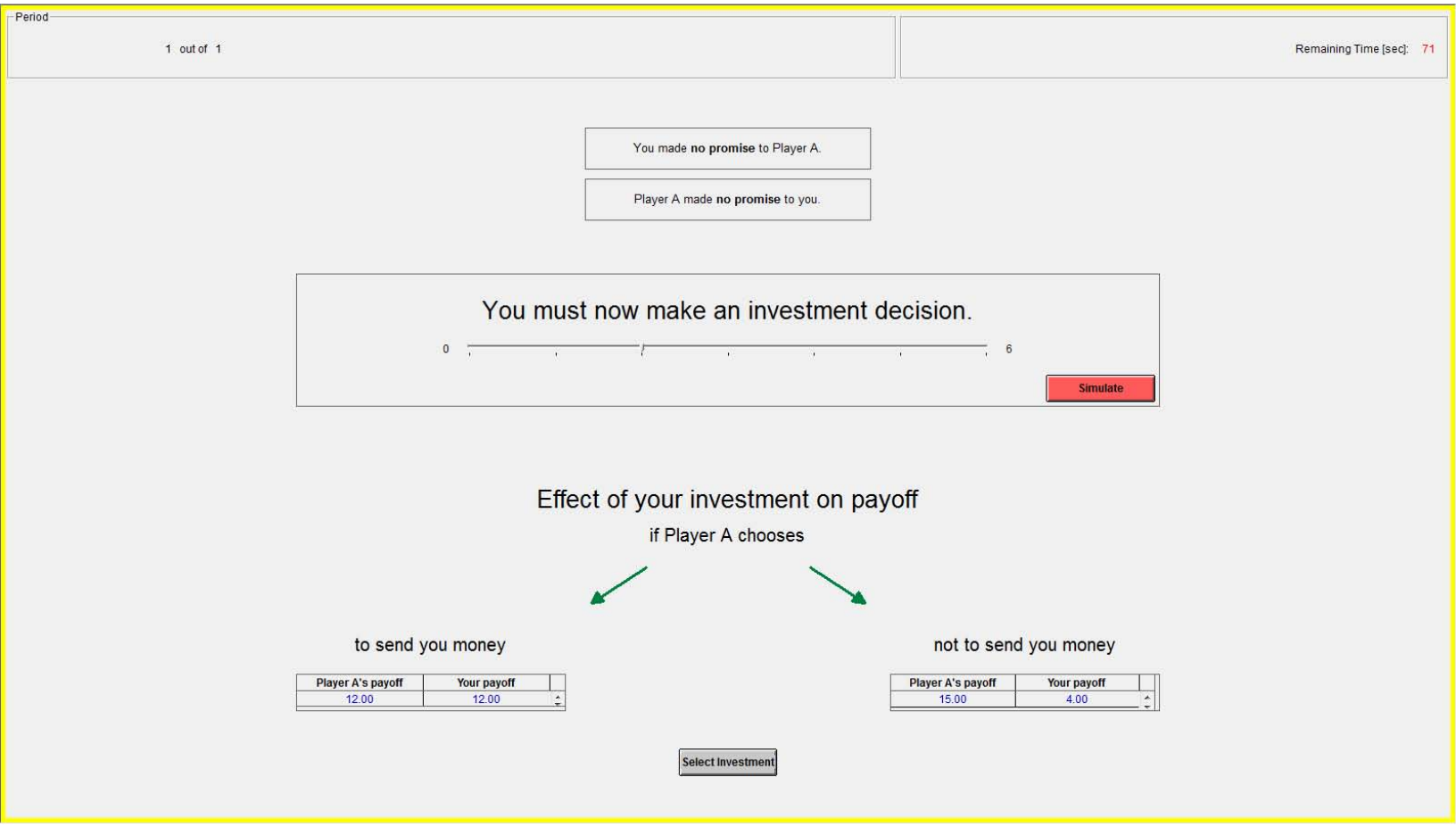
	Player A will certainly send you money	Player A will probably send you money	Unsure what Player A will do	Player A probably won't send you money	Player A certainly won't send you money
Your earnings if Player A sends you money	$0.65	$0.60	$0.50	$0.35	$0.15
Your earnings if Player A doesn't send you money	$0.15	$0.35	$0.50	$0.60	$0.65

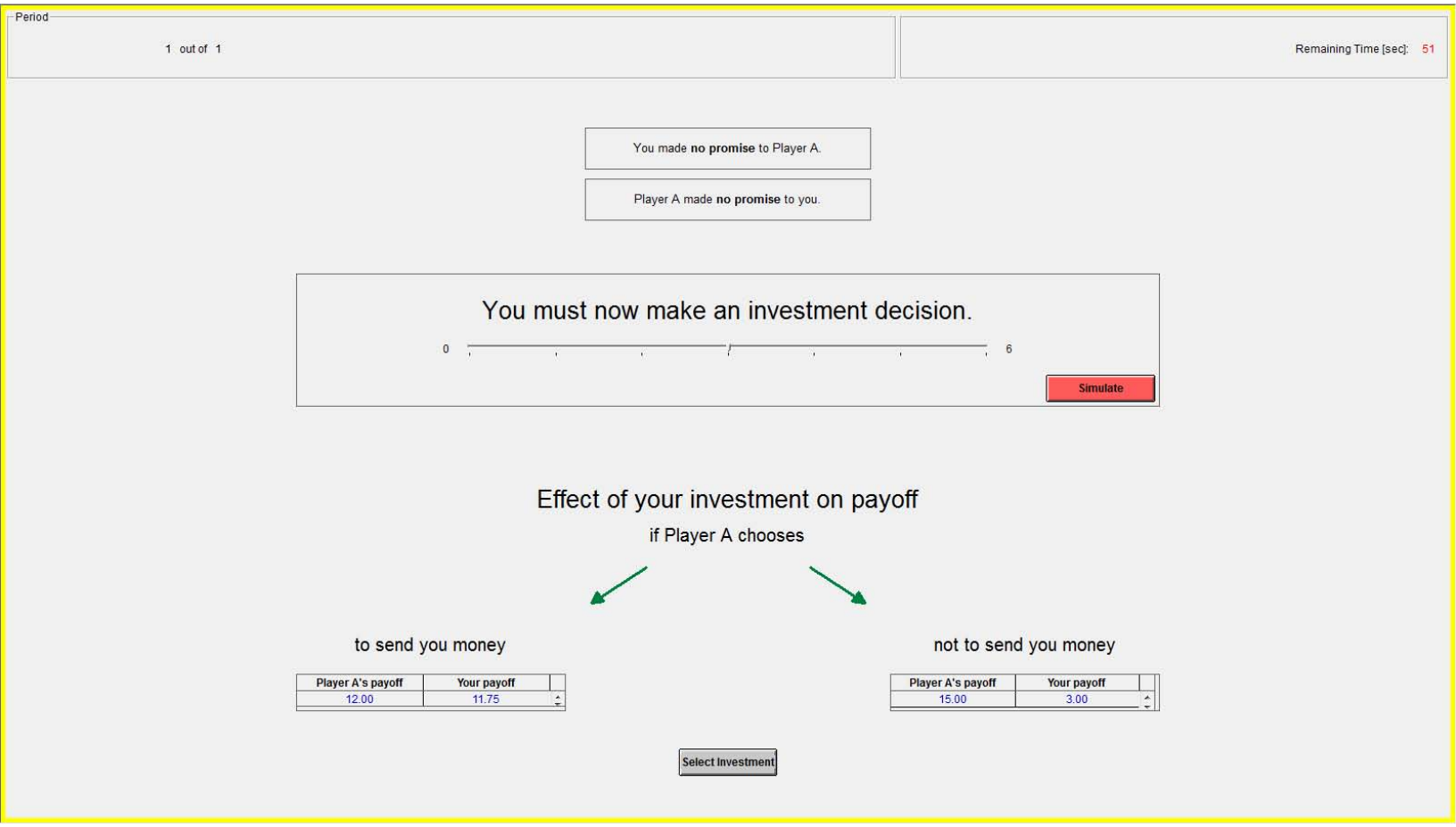
	Player A will certainly send you money	Player A will probably send you money	Unsure what Player A will do	Player A probably won't send you money	Player A certainly won't send you money
Investment is 0	○	○	○	○	○
Investment is 1	○	○	○	○	○
Investment is 2	○	○	○	○	○
Investment is 3	○	○	○	○	○
Investment is 4	○	○	○	○	○
Investment is 5	○	○	○	○	○
Investment is 6	○	○	○	○	○

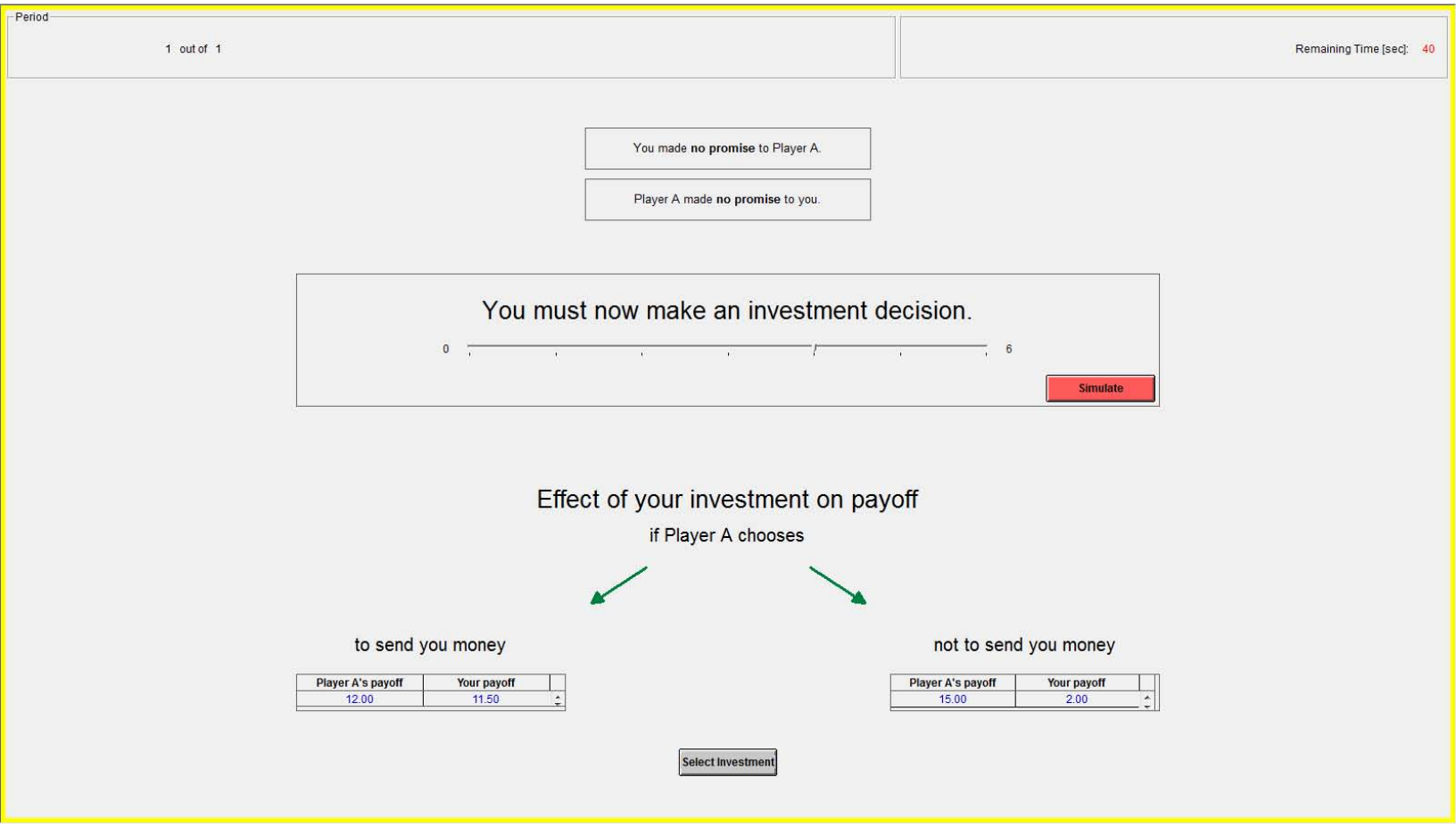
Continue

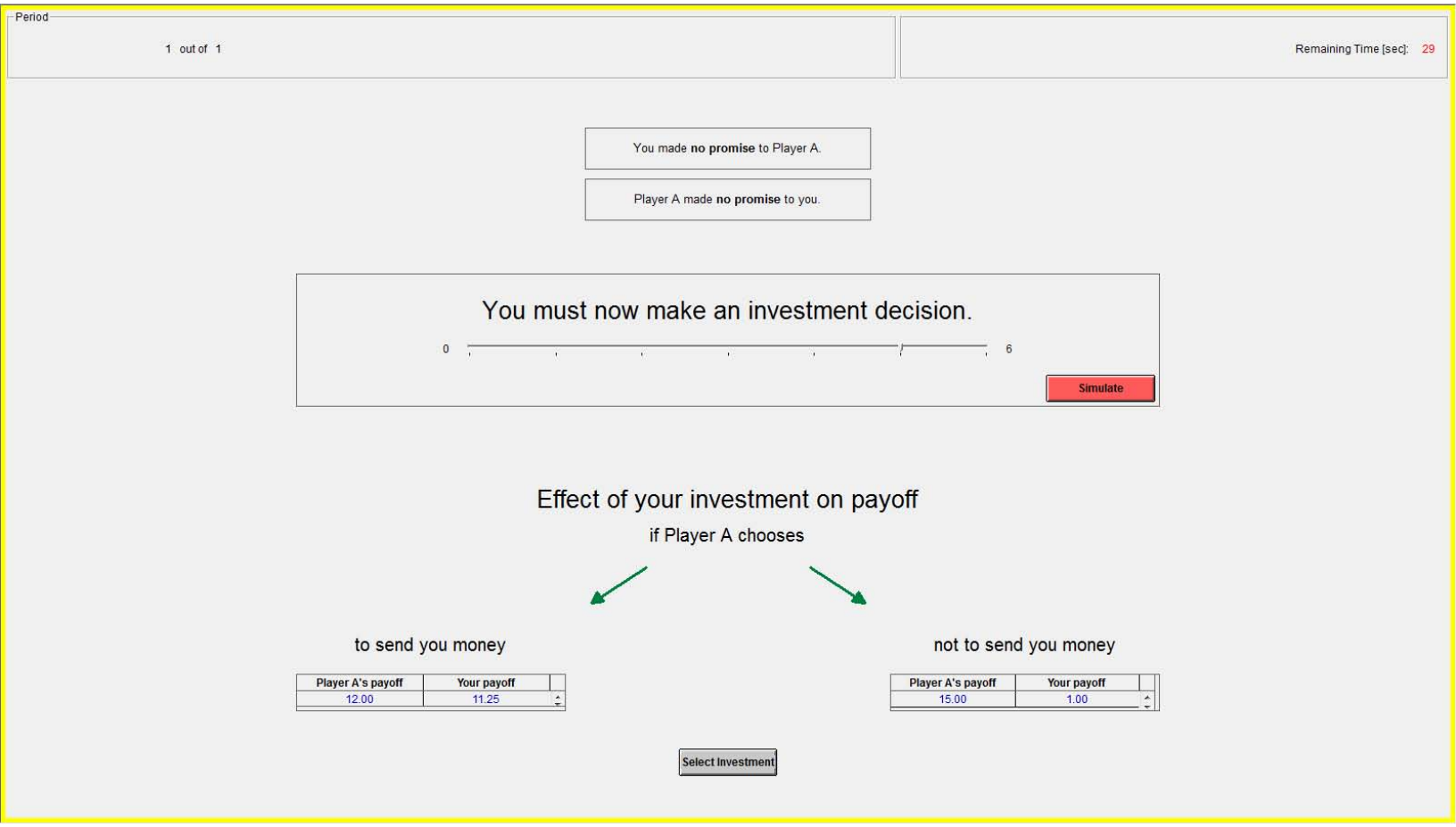


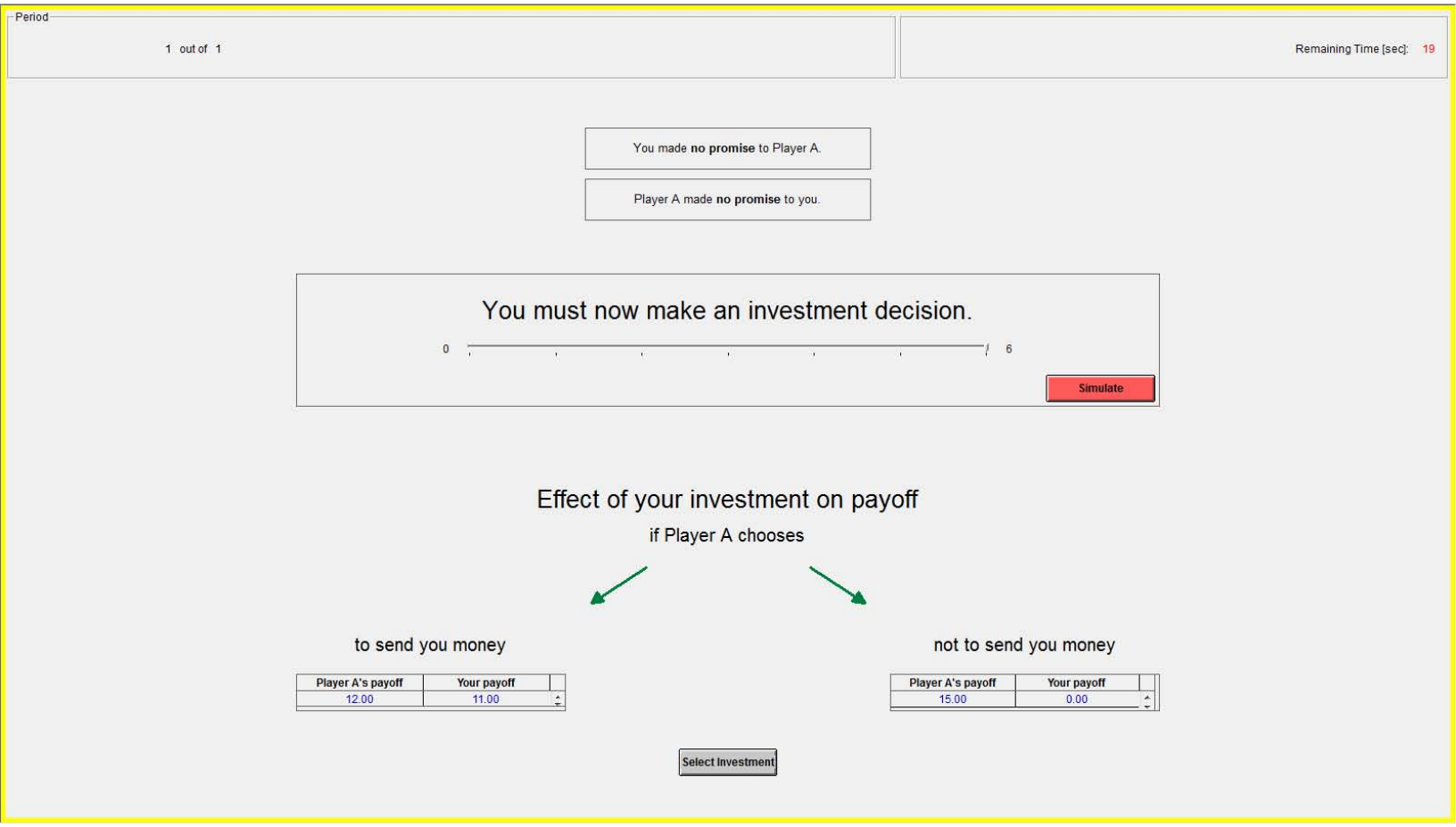


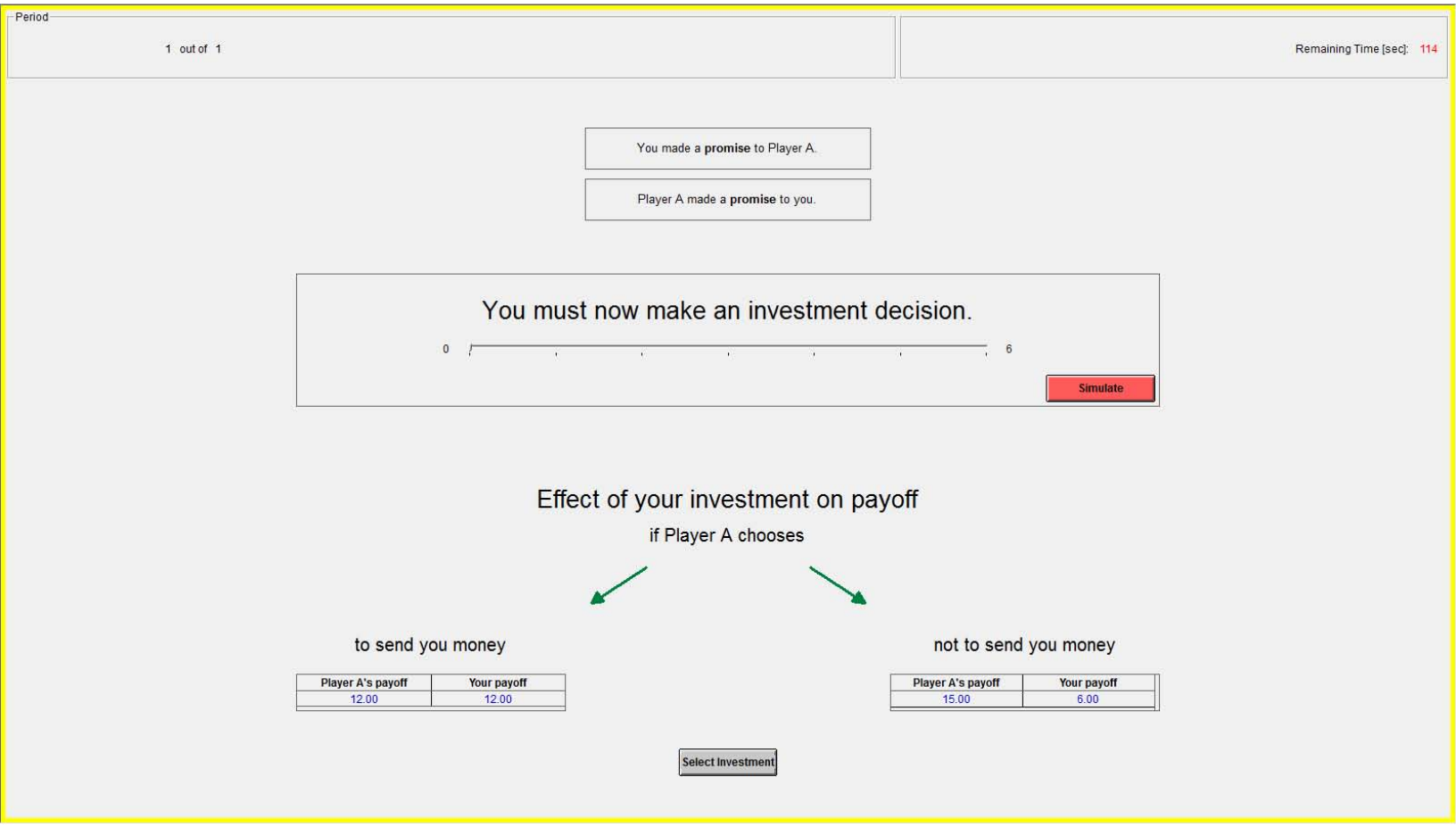


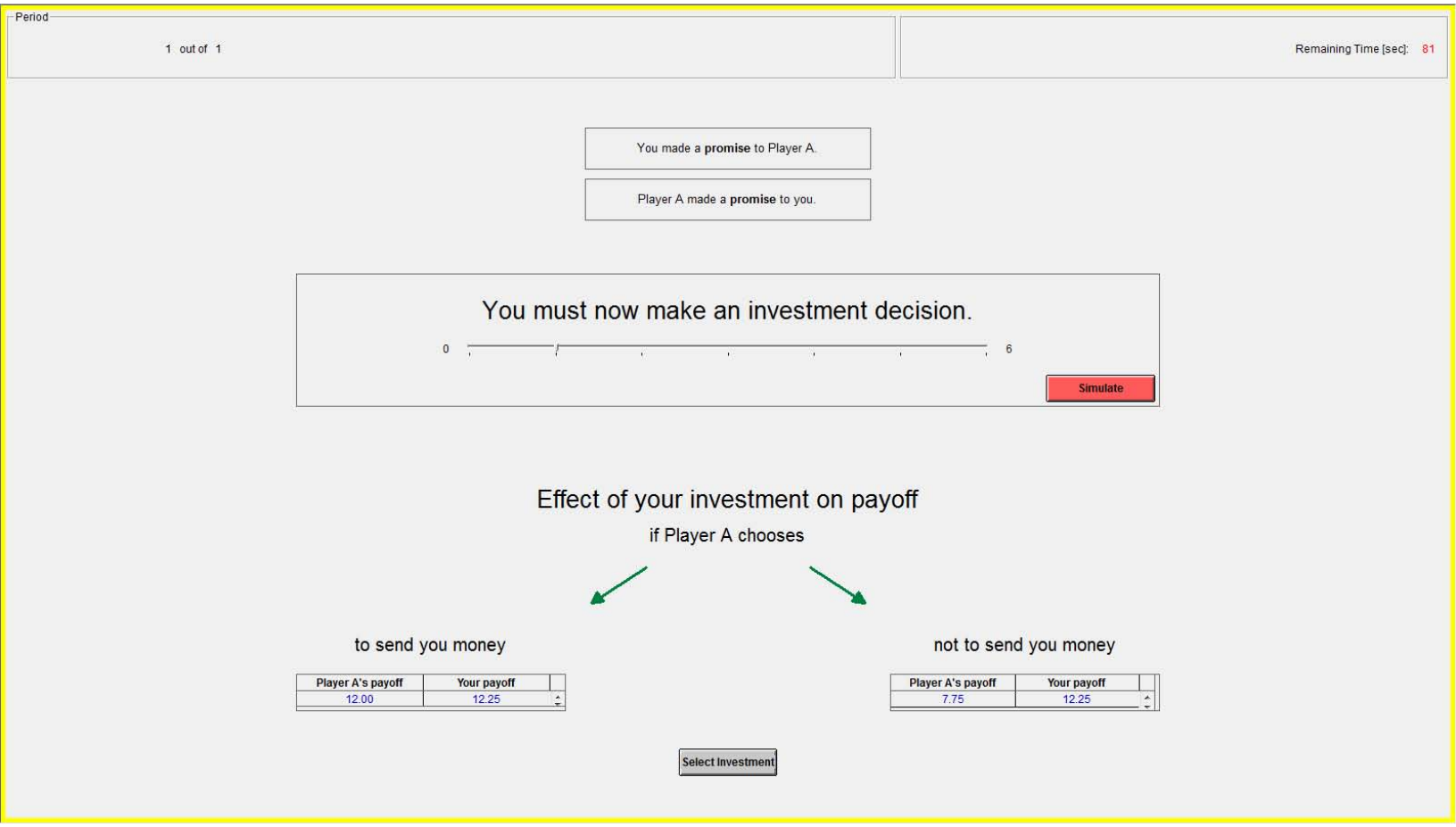


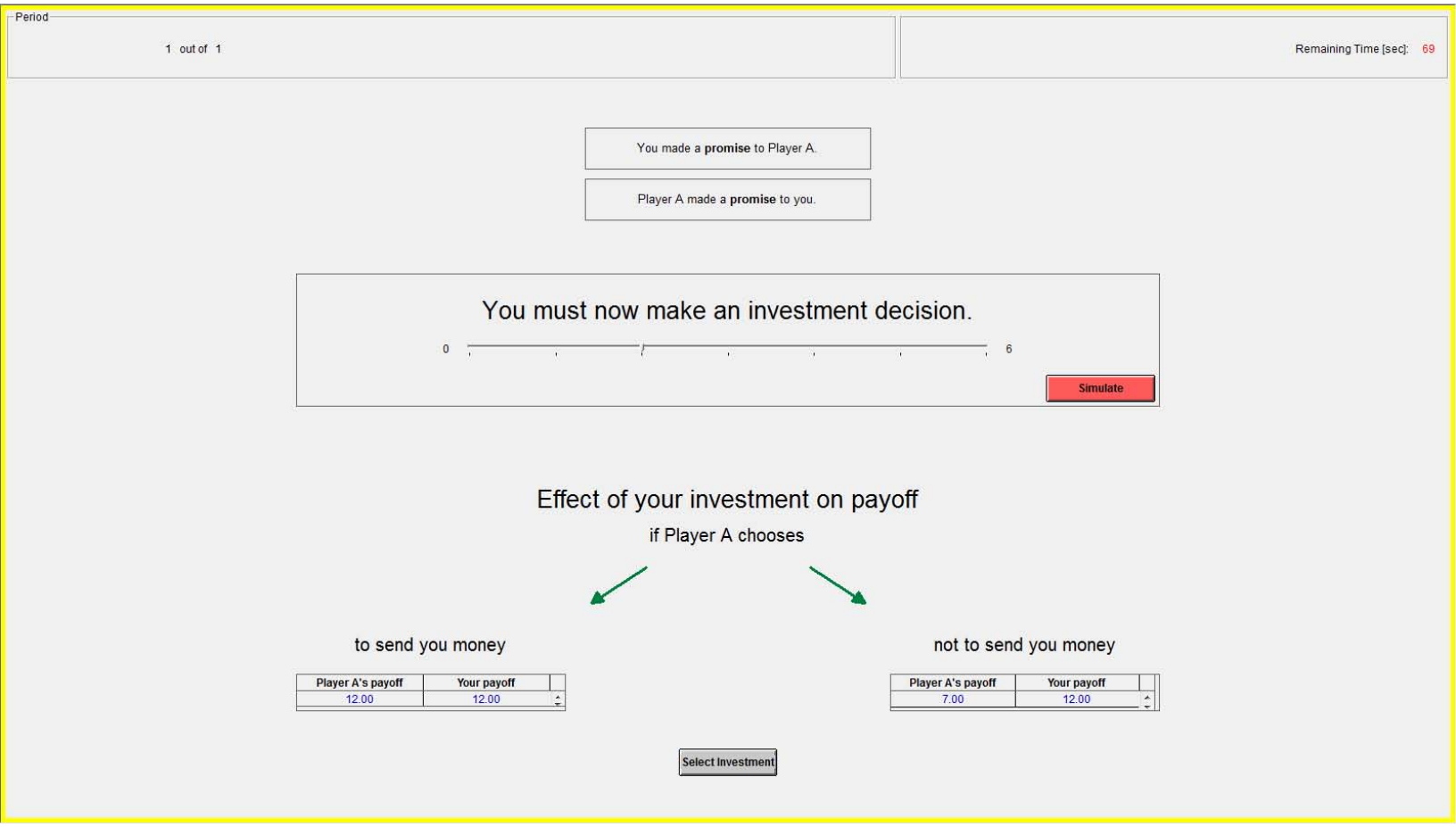


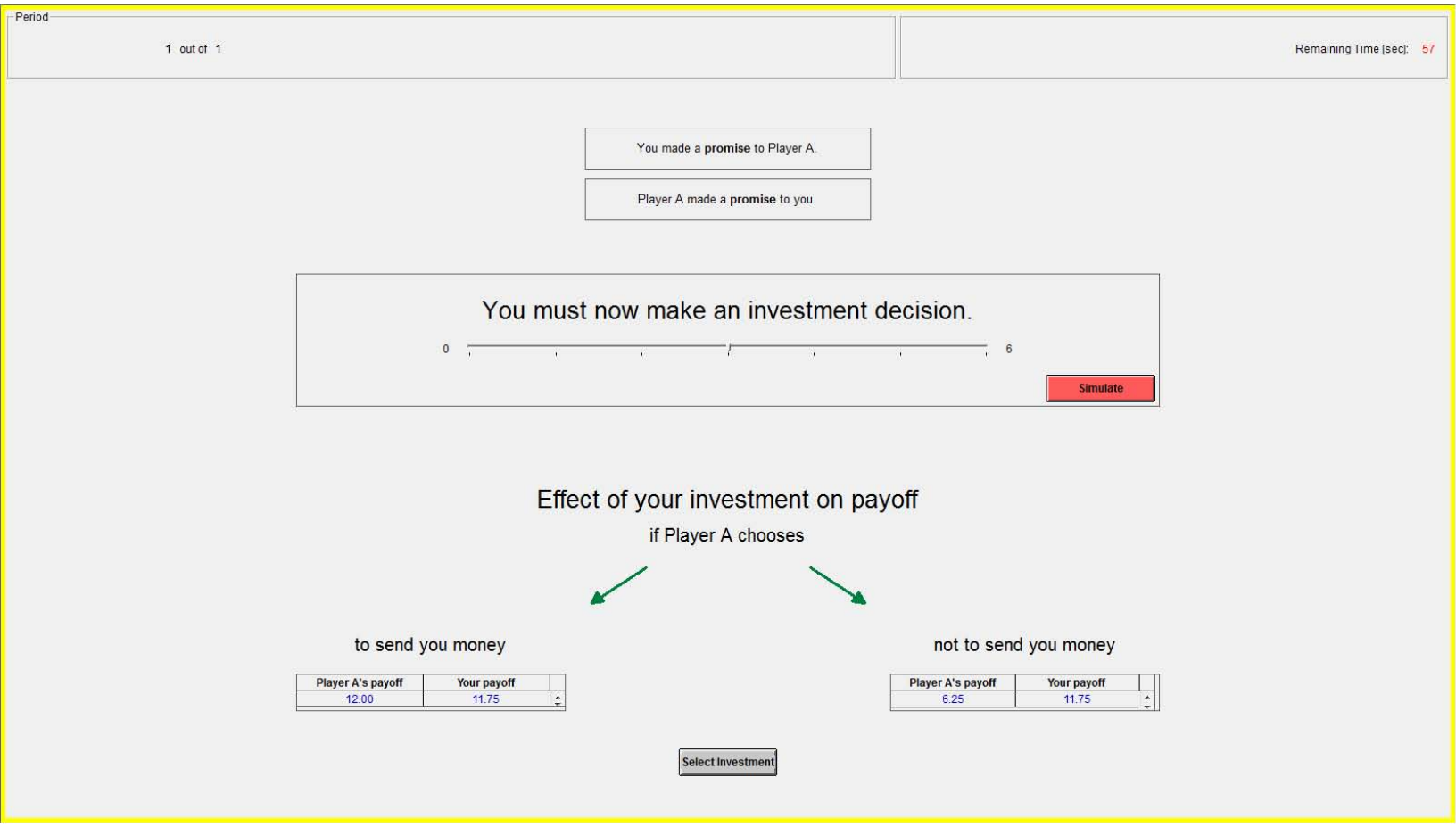


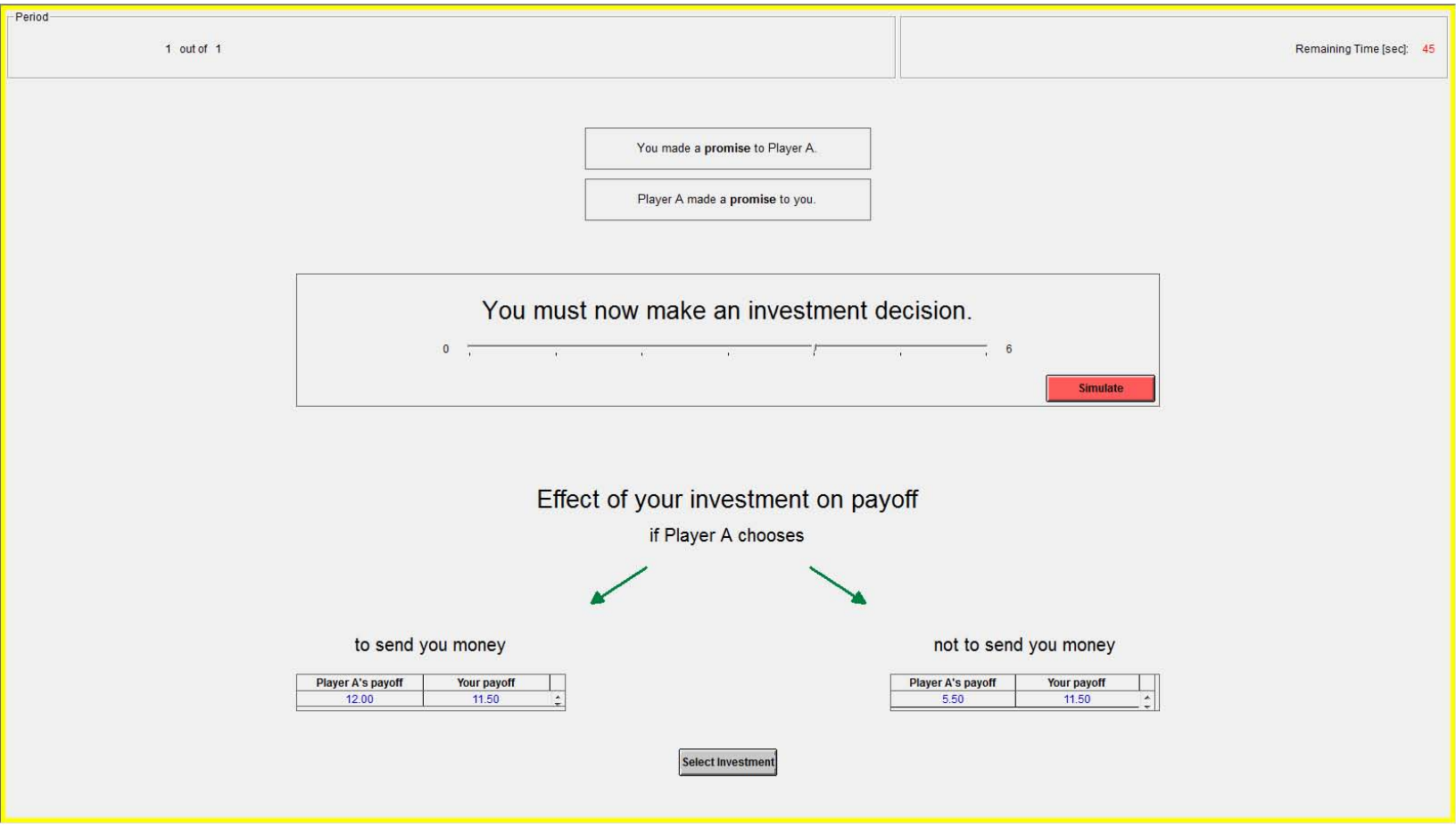


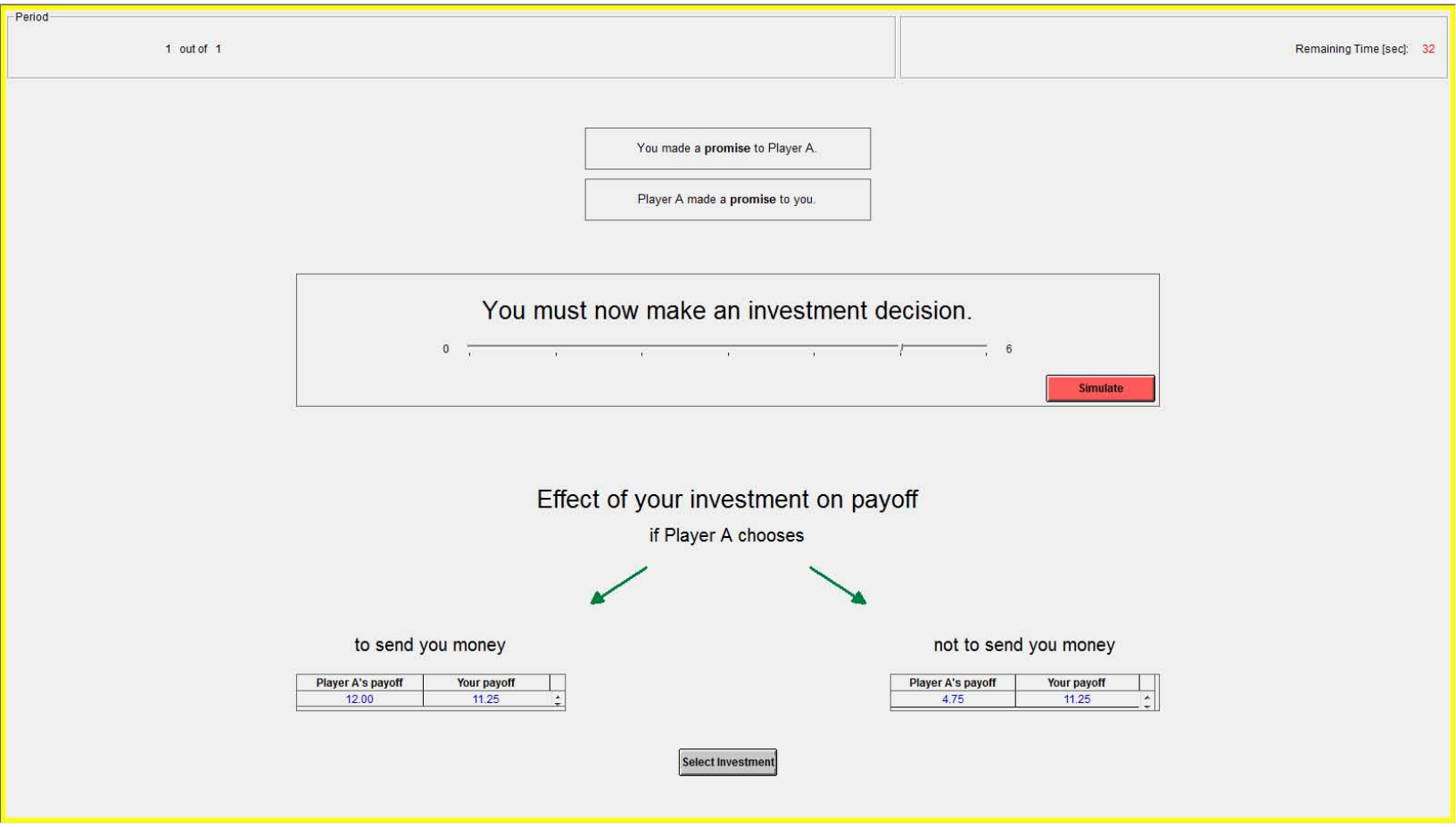


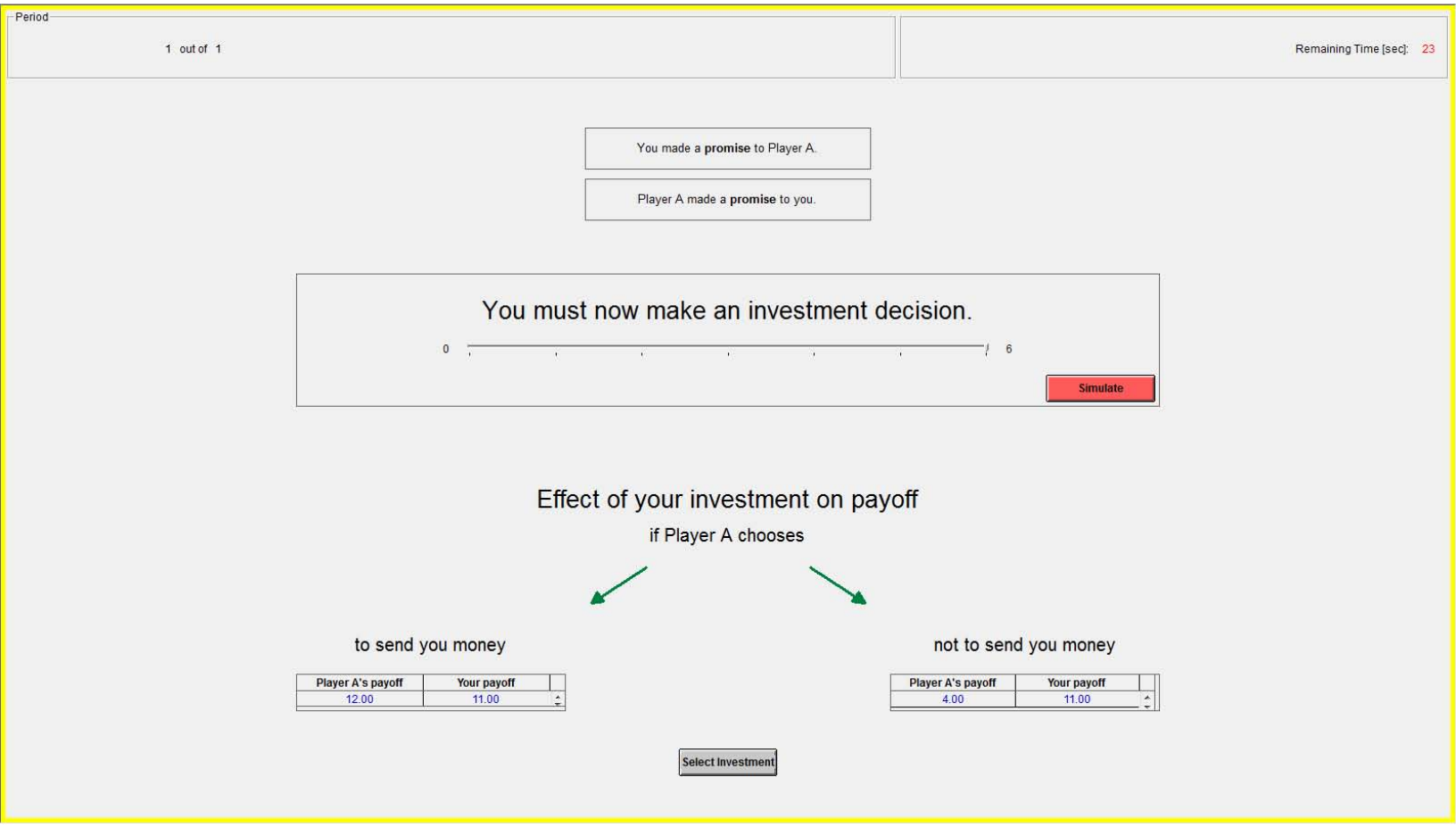












Period

1 out of 1

Remaining Time [sec]: 11

Your payoff is $12.00.

Continue

Period

1 out of 1

Remaining Time [sec]: 29

Your payoff is $11.00.

Your additional payoff from guessing is $3.50.

Continue

Period

1 out of 1

Remaining Time [sec]: 10

You are now progressing to a new round in which you will be randomly matched with a different participant.

Period

2 out of 2

Remaining Time [sec]: 29

Round 2 was chosen for payment based on both players' actions.

Round 1 was chosen for payment based on Player B's guesses.

You earned a total of $4.00.