Promises, Reliance, and Psychological Lock-In*

Rebecca Stone and Alexander Stremitzer†
UCLA

March 17, 2016

Abstract

In the absence of a legal regime enforcing promises, the classical prediction of contract theory is that promisees will underinvest in reliance whenever the promisor is likely to have a self-interested reason to break her promise. But if a promisor experiences guilt when she breaks her promise, this guilt may be intensified by the promisee's reliance on the promise. Anticipating this, the promisee has a strategic reason to overinvest in reliance in order to psychologically lock the promisor in to keeping her promise. A legal regime that enforces promises may therefore have the unexpected benefit of reducing overreliance as promisees no longer have to rely on the extra-legal mechanism of psychological lock-in in order to induce a promisor to keep her promise. We obtain experimental evidence supporting the existence of this psychological lock-in effect.

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

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

---

*The authors are grateful to Oren Bar-Gill, Gary Charness, Bob Cooter, Joseph Doherty, Gillian Hadfield, Rob MacCoun, Andreas Roider, Seana Shiffrin, and Kathy Spier. We are also grateful to seminar audiences at Berkeley, Boston University, Harvard, Paris II, UCLA, USC, the 2015 EMLS Workshop held at UCLA, and the ISNIE 2015 meeting. We thank Calvin Cam, Rita Hsu, Brandon Milostan, and Reggie Young for excellent research assistance. Financial Support by a 2014-15 UCLA Faculty Research Grant is gratefully acknowledged.

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

1 Introduction

In this paper, we experimentally investigate whether promisees overinvest in order to make their promisors more willing to keep their promises–an effect we refer to as "psychological lock in"–and whether legal enforcement of relied upon promises can therefore have the unexpected benefit of reducing overinvestment. We focus on three questions. First, does greater reliance by a promisee on a promise make the promisor more likely to keep the promise? Second, do promisees anticipate such an effect, and so strategically increase their reliance on promises in order to make their promisors more likely to keep them? Third, what benefits for social welfare arise from legal enforcement of promises that have been relied upon? In particular, in what ways does such legal enforcement improve promisees' investment decisions? And does it improve rates of promise keeping?

In the absence of legal enforcement, a promisee may be wary of relying on a promise for fear that the promisor won't keep her promise. Since classical economic theory assumes that agents are rational and self-interested, it predicts that a promisee won't be willing to invest very much 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 therefore 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. Of course, if, contrary to the predictions of classical theory, many promisors are intrinsically motivated to keep their promises, this underinvestment will be reduced even in the absence of a legal regime. But since many people behave opportunistically some of the time, departures from the self-interest assumption are unlikely to eliminate the problem entirely. A large literature on breach remedies and the holdup problem studies how the introduction of third-party enforcement can mitigate this underinvestment problem.[1]

But reducing underinvestment might not be the only virtue of legal enforcement if many 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).[2] Suppose further, as our model assumes, that a promisor's intrinsic motivation to keep a promise increases the more the promisee has relied upon it, even when such reliance does not confer any material benefit on the promisor.[3] If that is the case, then the promisee has an incentive to strategically rely on the promise in order to make the promisor more likely to keep it. The result is that the promisee might *overinvest* rather than underinvest in the absence of a legal regime in order to psychologically lock in the promisor. Thus, legal enforcement of relied-upon promises may reduce a promisee's felt need to overinvest in reliance on a promise, as the promisee no longer has to rely on this extralegal mechanism of

[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]For notable contributions to the broader literature in political science and social psychology, see Ostrom, Walker, and Gardner (1992), Kerr and Kaufman-Gilliland (1994), Sally (1995), and Bicchieri and Lev-On (2007)

[3]Consistent with this conjecture, Wilkinson-Ryan and Hoffman (2015) provide suggestive evidence based on a psychological vignette study that people are more likely to keep relied upon promises, though in their study the reliance confers a material benefit on the promisor.

psychological lock-in. In other words, legal enforcement may have the unexpected benefit of reducing overinvestment alongside the more expected benefit of reducing underinvestment and increasing rates of promise keeping.

The form of legal regime that we study in this paper mimics a legal regime that is governed by the doctrine of promissory estoppel. Pursuant to this doctrine, courts enforce gratuitous promises–promises for which the promisor received nothing in return–but only when they have been relied upon. The doctrine is also invoked to enforce promises that have been relied upon that are part of a bargained-for exchange, and therefore presumptively enforceable even in the absence of reliance, when those promises would be rendered unenforceable for some other reason such as lack of definiteness or a failure to satisfy the Statute of Frauds.[4] Judges have discretion to determine the remedy when they find a promisor liable on promissory estoppel grounds. In particular, they can choose between Expectation Damages–the standard remedy for breach of contract–and Reliance Damages.[5] But the scholarly consensus seems to be that Expectation Damages are more commonly awarded (Farber and Matheson, 1985; Yorio and Thel, 1991). Thus, in this paper we study the effects of introducing a legal regime with Expectation Damages by contrasting it with a regime in which legal enforcement is absent.

Our subjects played a modified dictator game in which a Recipient first makes an investment decision and then a Dictator decides whether or not to "cooperate" with the Recipient. Cooperation always increases joint payoffs, but, in the absence of a legal regime, it reduces the Dictator's payoff. Investment by the Recipient affects

[4] See, e.g., Jamestown Terminal Elevator, Inc. v. Hieb, 246 N.W. 2d 736 (N.D. 1976).
[5] See Restatement (Second) of Contracts § 90 (1981).

only his own payoffs in the absence of a legal regime. The way in which it does so depends on whether the Dictator cooperates with him. If the Dictator chooses the uncooperative action, investment monotonically reduces the Recipient's payoff. If the Dictator decides to cooperate, then investment first increases and then decreases the Recipient's payoff. Thus, the payoff-maximizing level of investment is positive, but only if the Dictator cooperates.

Before subjects learned their roles in this game, each had the opportunity to make a promise to the other to cooperate in the event that he was chosen to be the Dictator. In our No Regime treatment, the Dictator suffered no penalty if she broke a promise to cooperate with the Recipient. In our Expectation Damages treatment, she suffered such a penalty if the Recipient had invested in reliance on this promise, as she was forced to pay the Recipient expectation damages.

Consistent with the predictions of our model, we find that the incidence of overinvestment is higher under the No Regime treatment than under the Expectation Damages treatment. We also find that those who overinvest overinvest more under No Regime than under Expectation damages. Finally, we find that promisees believe that promisors are likely to cooperate the more those promises were relied upon. These findings provide powerful evidence for the presence of a psychological lock-in effect.

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 word in order to avoid failing to meet the expectations of performance that their promises create in their promisees. A promisor is therefore more likely to keep her promise if she believes that the promisee expects her to keep her promise (Dufwenberg and Gneezy, 2000; and Charness and Dufwenberg, 2006). In contrast, the *commitment-based* theory claims that promisors have a preference for keeping their word independent of the expectations of promisees (Braver, 1995;, Ostrom, Walker, and Gardner, 1992; Ellingsen and Johannesson, 2004; Vanberg, 2008; and Ismayilov and Potters, 2012).

It is, of course, possible that *both* the promisee's expectations of performance and a promisor's sense that she is duty-bound to keep her promise regardless of the promisee's expectations contribute to her willingness to keep her promise. Vanberg (2008), using an ingenious experimental design, argues that this sense of duty is the *only* reason why people keep their promises. But his suggestion that the promisee's expectations are irrelevant to promise keeping is based on the observation that exogenous variation of second-order expectations–that is, variation in promisors' beliefs about promisees' beliefs–doesn't lead to any statistically significant difference in promise keeping in a setting where there are no direct promissory links between the parties. In Vanberg's design, subjects were rematched after a round of promises has been made, so that promisors faced promisees to whom they made no 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. Here too, direct promissory links between the subjects were missing because no promises were ever made.

In Charness's and Dufwenberg's (2006) experiment, by contrast, there were promissory links between subjects, and the results provide suggestive evidence that expectations matter. But their evidence is based on correlations between second-order beliefs and actions that also admit alternative explanations. Ederer and Stremitzer (2015) get around this problem by exploiting exogenous variation in promisors' second-order beliefs. They show that the irrelevance of promisor expectations is a unique feature of the special cases in which expectations are not supported by direct promises. Instead, promise keeping seems to have a lexicographic structure: a promisee's expectations matter if and only if the promisor actually made a promise to the promisee, or at least, the presence of a promise reinforces promisors' sensitivity to promisees' expectations.

We assume that a promisor's motivation to keep her promise increases as the promisee's reliance upon the promise increases. If promisors don't like defeating another's expectations, it is plausible to suppose that they also don't like disappointing promisees who have relied on their promises. This might be because a promisee's reliance is an indicator of his expectations that the promisor will perform. But it also might be because the promisor simply doesn't like to make the promisee worse off than he would have been had the promise been kept. Disappointed expectations are a kind of psychic harm that can result from the breaking of a promise. Reliance similarly results in harm–material rather than psychic–if a promise is broken, and greater reliance reduces the promisee's payoff further below the payoff he would receive if the promise were kept. Our formulation of preferences is designed to capture a concern with avoiding the latter kind of reliance-based harm. We don't suppose that promisors are influenced by their beliefs about their promisees' expectations, though we don't mean to suggest that such beliefs don't matter. We simply posit that a promisor exhibits a stronger preference to keep her promise the greater is the difference between the payoff the promisee would have received had the promise been kept and the payoff he would receive if the promise was broken. This implies that the more the promisee relied upon the promise, the more willing the promisor will be to keep it. In other words, although our model has a flavor of the expectation-based approach, it does not assume that promise-keeping is driven by the promisor's second-order beliefs. So, unlike the expectation theorists, we don't need to employ the apparatus of Battigalli's and Dufwenberg's (2007, 2009) psychological game theoretic model according to which agents experience "guilt" when their behavior falls short of another's expectations.[6]

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 shows that, consistent with the predictions of classical economic theory (e.g., Shavell, 1980), 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 promise is inefficient. The

[6] We do make use of their terminology of "guilt." But the idea of "guilt aversion" need not be taken too literally. What matters for our purposes is that the promisor exhibits a greater willingness to keep her promise the more she believes that the promisee has relied upon it. This could be because the psychological sensation of guilt from breaking the promise is increasing in the promisee's reliance. But it could also be because she simply believes that it is a graver moral wrong to break a promise that has been relied upon more, regardless of any guilt she might experience were she to break it.

promisee disregards this possibility when choosing his investment level and so invests the amount that would be desirable if performance were always efficient.[7] Our design abstracts from this motive to overinvest by ensuring that promise-keeping is always efficient.

Subjects don't make promises in the experiments of Sloof et al. (2003) and Sloof et al. (2006). The contracts that they are imagined to have signed are simply given to them at the outset of the game. In our design, by contrast, subjects can decide whether or not to make promises to one another and so promises arise endogenously. We allowed subjects to decide whether or not to make promises to one another because we wanted the promises in the experiment to be perceived to be genuine promises. In this respect, our design more closely resembles that of Hoppe and Schmitz (2011), who find that option contracts improve investment incentives even when the option contract is not enforceable. The authors assume that these option contracts only arise if accepted by both parties, and so their agreements are endogenous much like our promises are. Their focus, however, is different from ours.

The remainder of the paper is organized as follows. Section 2 presents the model from which we derive our theoretical predictions. Section 3 sets out our experimental design and formulates the specific hypotheses that we test. Section 4 presents our results, which are largely in line with our theoretical predictions. Section 5 offers a further discussion of our results. Section 6 summarizes and concludes.

[7] 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, under reliance damages, the investor has an additional incentive to invest to reduce the likelihood of breach.

2 Theory

In this section, we develop the theoretical model that we use to generate the hypotheses that we test using the data from our experiment.

2.1 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. At the outset of the game, the Dictator decides whether to make a promise to cooperate with the Recipient, $p \in \{0, 1\}$. This promise has no effect on either agent's material payoffs in the absence of legal enforcement. The Recipient then chooses an investment level $i \in [0, \bar{i}]$. Finally, after observing the Recipient's choice, the Dictator chooses an action $a \in \{0, 1\}$ where $a = 1$ denotes the cooperative action.

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

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

and therefore

$$\pi_R(1, i) > \pi_R(0, i). \tag{2}$$

Assumption 2. In the absence of a legal regime, investment affects only the Recipient's payoff. Thus, the Dictator would never have a reciprocity-based reason to cooperate the larger was the Recipient's investment, and so we abstract from such motives when we model the Dictator's preferences in Section 2.2 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:

$$\pi'_R(0, i) < 0, \tag{3}$$

while

$$\pi'_R(1, i) > 0 \text{ if } i \leq e, \tag{4}$$

and

$$\pi'_R(0, i) < \pi'_R(1, i) < 0 \text{ if } i > e, \tag{5}$$

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:

$$\pi'_R(1, i) - \pi'_R(0, i) > 0. \tag{6}$$

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.[8] Consistent with Assumption 1 we assume that the Dictator's material payoff is given by:

$$\pi_D(a,i) = \begin{cases} 15 & \text{if } a = 0 \\ 12 & \text{if } a = 1 \end{cases}, \tag{7}$$

and the Recipient's payoff is given by:

$$\pi_R(a,i) = \begin{cases} 6 - i & \text{if } a = 0 \\ 12.25 - 0.25\,|1-i| & \text{if } a = 1 \end{cases}, \tag{8}$$

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.

2.2 Preferences

Because the Recipient makes no promise, 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 given simply by his own material payoff:

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

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

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 in the event that she made a promise to cooperate with the Recipient–a desire that is intensified, the more the Recipient has to lose if the Dictator breaks the promise. We therefore posit that the Dictator's utility is reduced if she breaks a promise by an amount that depends on the reduction of the Recipient's material payoff that results from the broken promise. That is, her utility depends on her material payoff, and, if she made such a promise, the difference between the Recipient's actual material payoff and the material payoff the Recipient would have received if she had kept her promise.[9] In line with the economics literature on promising, we refer to latter component of the utility as the Dictator's "guilt." (Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007, 2009; Ederer and Stremitzer, 2015). Formally, the Dictator's utility is given by:

$$u_D(a,p) = \pi_D(a) - p\gamma g\left(\pi_R(1,i) - \pi_R(a,i)\right), \tag{9}$$

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 (6) implies that guilt from not keeping a promise increases in investment, so $\frac{dg}{di} > 0$ if $a = 0$.

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

[9] 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 (see Ederer and Stremitzer, 2015) or a simple desire to do as one promised.

use of the term "guilt" might be taken to suggest that the Dictator's psychological well-being is reduced when she breaks a promise. On this literal interpretation, the "guilt" function describes a self-interested motivation to keep a promise–the motivation to avoid an unpleasant emotional state–and so ought to be regarded as a component of the Dictator's welfare and therefore included in the social welfare function, W. But we need not view the "guilt" function in this way. We could instead regard it simply as capturing the strength of the non-self-interested considerations that the Dictator perceives give her reasons to keep her promises. That is, it might simply capture the fact that the Dictator believes that the moral reasons to keep a promise are more weighty the larger the reduction in payoff that the promisee would suffer were she to break the promise, regardless of any effect that breaking her promise has on her psychological well-being. In that case, "guilt" would not count as a component of her welfare, and so ought not to be included in the social welfare function.

2.3 Complete Information Equilibrium without Legal Enforcement

We now solve the game by backward 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. For the moment, we ignore the promise-making stage of the game and so take the Dictator's decision whether to make a promise as exogenous. We explain why we expect to see promises in equilibrium in Section 3 when we introduce the promise-generating technology that we use in the experiment.

2.3.1 Equilibrium without Guilt

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, in the absence of guilt, 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 (3). If the Dictator chooses to cooperate, then expressions (4) and (5) imply that the recipient will choose e. But expression (1) implies that the Dictator's dominant strategy is not to cooperate. ■

2.3.2 Equilibrium with Guilt

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 9 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, since $\pi_R(0, i)$ is strictly decreasing in i. If, by contrast, the Dictator made a promise, then the Dictator will cooperate whenever

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

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

$$\Leftrightarrow \gamma \geq \gamma^c(i) = \frac{\pi_D(0) - \pi_D(1)}{g(\pi_R(1, i) - \pi_R(0, i))}. \tag{10}$$

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 experienced from breaking 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 (6) that increasing investment i increases $\pi_R(1,i) - \pi_R(0,i)$ and so, given (10), 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. See Appendix A. ■

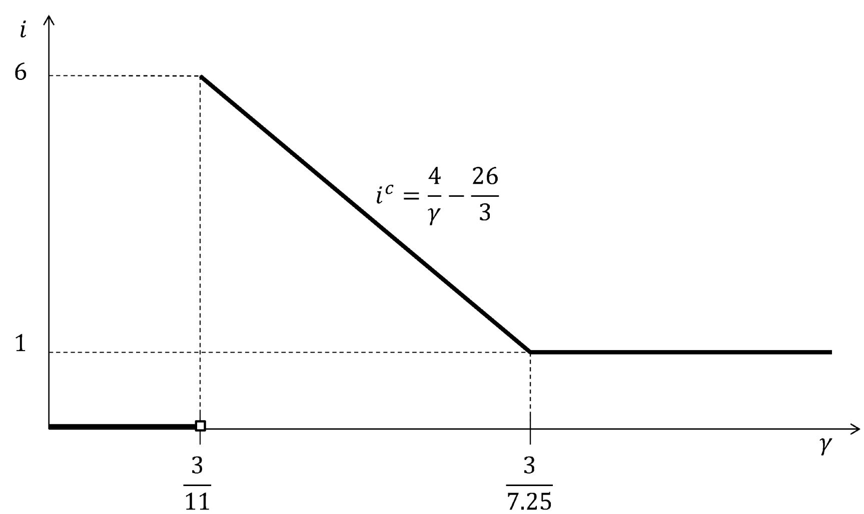

Figure 1: Equilibrium investment as function of γ

Figure 1 illustrates Proposition 4 assuming a linear guilt function and the specification of the payoff functions that we used in our experiment. We see from Propositions 3 and 4 that once we introduce guilt into the Dictator's utility function, there are several possibilities depending on the size of the guilt parameter. If the guilt parameter is so low that the Dictator wouldn't cooperate even if the Recipient chose the maximum investment level, then the Recipient won't invest and the Dictator won't cooperate. The resulting equilibrium is indistinguishable from the equilibrium that results when Dictators only care about their material payoffs, and the agents fail to achieve the first-best. If the Dictator's guilt parameter is high enough that investing the efficient amount is sufficient to induce the Dictator to cooperate, then the Recipient will invest the efficient amount and the Dictator will cooperate. The result is that the equilibrium outcome maximizes social welfare. Finally, if the Dictator's guilt parameter is in an intermediate range, the Recipient will invest more than the efficient level by an amount that is sufficient to induce the Dictator to cooperate.[10] Thus, guilt can also produce a new kind of inefficiency: overinvestment motivated by the Recipient's desire to psychologically lock the Dictator into keeping her promise.

2.4 Complete Information Equilibrium 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

[10] This result only obtains if the Recipient's payoff when he invests the amount required to induce the Dictator to cooperate is not reduced below the payoff he would receive were he to invest zero. If this condition does not obtain, then the Recipient chooses a zero investment level and the Dictator won't cooperate.

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:

$$\pi_D^L(0,i) = \pi_D(0) - l(i) = \begin{cases} \pi_D(0) & \text{if } i = 0 \\ \pi_D(0) - l(i) & \text{if } i > 0 \end{cases}, \tag{11}$$

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:

$$\pi_D^L(1,i) = \pi_D(1), \tag{12}$$

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, $i = 1$.[11] 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 *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 will not invest and the Dictator will not cooperate, $\left(\widehat{a}_L, \widehat{i}_L\right) = (0, 0)$. (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 made 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 therefore be 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

[11] 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, 84), Cooter and Eisenberg (1985) and others, expectation damages won't produce overinvestment in our model.

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 allows the agents to achieve the first best when a promise was made.

2.5 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:

$$\begin{aligned} U_R(i) &= E\left[\pi_R(a,i)\right] \qquad (13) \\ &= P\left(\gamma < \gamma^c(i)\right)\pi_R(0,i) + \left(1 - P\left(\gamma < \gamma^c(i)\right)\right)\pi_R(1,i) \\ &= \pi_R(1,i) - P\left(\gamma < \gamma^c(i)\right)\left(\pi_R(1,i) - \pi_R(0,i)\right), \end{aligned}$$

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.[12] 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.*

In Appendix A, we work through a specific example with three Dictator types: a low-guilt type who never cooperates; an intermediate-guilt type who cooperates if the Recipient overinvests; and a high-guilt type who cooperates regardless of the Recipient's investment level. Using the payoff specifications that we use in the experiment and a linear guilt function, we show that the Recipient will optimally overinvest so long as the likelihood that the Dictator is of an intermediate-type is sufficiently high. By contrast, uncertainty about the Dictator's type has no effect on the equilibrium in the expectation damages regime. This is because the legal regime ensures

[12] In Appendix A, we explore the case of uncertainty over Dictator types by using a simple example.

that no matter what the Dictator's type, the Recipient can ensure that the Dictator cooperates by choosing the first-best investment so long as the Dictator made a promise.

In the rest of the paper, we put our theory to an experimental test.

3 Experimental Design & Procedure

3.1 Design

In each round of the experiment, each subject was first randomly and anonymously matched with another subject. The subject pairs then played a modified dictator game that resembles the game described in Section 2 and is depicted in Figure 2.

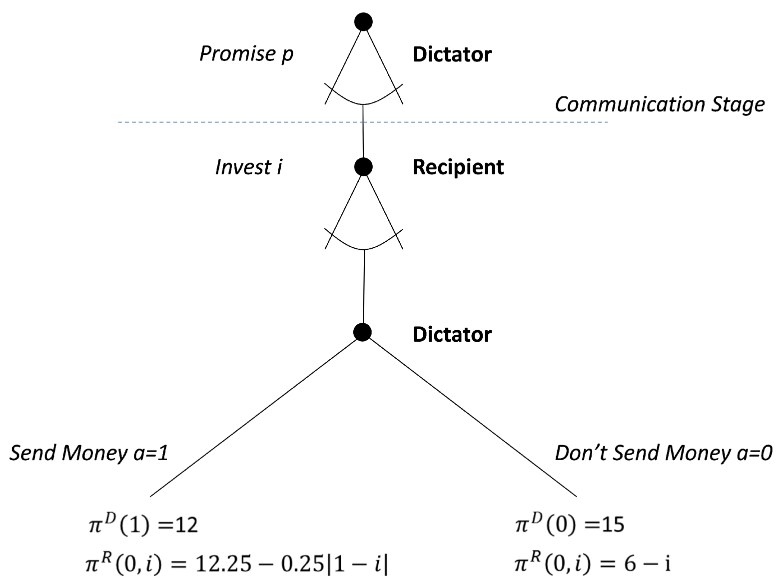

Figure 2: Game Tree No Regime.

Two features of our design differentiate it from a standard dictator game. First, before subjects learned their role as either the Dictator ("Player A") or the Recipient

("Player B"), each had the opportunity to promise to send the Recipient money (the "cooperative" action) in the event that she was chosen to be the Dictator. Second, after subjects had learned their role for the round, the Recipient had to make his investment decision before the Dictator decided whether or not to send the Recipient money.

As we noted when describing Assumption 2 in Section 2.2, the investment decision affected only the Recipient's payoffs in the No Regime treatment. Thus, the Dictator had no reciprocity-based reason to reward the Recipient for a positive investment choice. This was important because our aim was to isolate the effects of reliance alone on promise keeping. In both treatments, basic material payoffs in the absence of legal enforcement were given by (7) and (8). The only difference was that the Recipient's choice set was discrete rather than continuous: $i \in \{0, 1, 2, 3, 4, 5, 6\}$.

In all treatments, we used the strategy method to elicit the Dictator's cooperation decision (Selten, 1967). The Dictator made her choice in ignorance of the Recipient's investment choice by indicating whether or not she wanted to cooperate for each possible investment level that the Recipient might have chosen. That is, each of the Dictator's choices is contingent on a particular investment level being chosen by the Recipient. The Dictator's actual choice (and therefore the players' final payoffs) was then determined by the investment level that the Recipient actually chose. If, for example, the Recipient chose to invest three, then the Dictator's action was the action that she indicated she wanted to choose in the event that the Recipient chose an investment of three.

We also elicited the Recipient's beliefs about the choices the Dictator made. Before the Recipient made his investment decision, we asked him to indicate how confident he was that the Dictator would choose to cooperate with him for each of the available investment levels.

During the communication stage, subjects exchanged computer-coded messages. One randomly-selected subject in each pair (Participant 1) first had to decide whether to promise to send money to the other subject (Participant 2) in the event he was chosen to be the Dictator. This promise was conditional on Participant 2 making a return promise.[13] Participant 2 could then choose whether or not to make a return promise to send money to Participant 1 in the event he was chosen to be a Dictator.[14] Subjects had reasons to make promises because the communication stage took place before they learned their roles. In No Regime, promising was cheap talk in the sense that there was no monetary cost associated with making a promise. However, there was a potential benefit if making a promise was interpreted as a kind act that subjects then reciprocated with kindness by cooperating if they ended up in the role of Dictator. In Expectation Damages, promising was no longer cheap talk, but making a promise was likely to yield a promise in return, which was beneficial for subjects in the event they ended up in the role of Recipient.

Alternatively, we could have used a version of the trust game that resembles our modified Dictator game, except that it gives the Recipient the opportunity to opt out of the game at the outset and the Dictator alone the opportunity of making a promise to cooperate with the Recipient at the outset of the game after the subjects' roles are revealed to them. Opting out gives both parties their outside options, which

[13] Participant 1's alternative to making a conditional promise was to send an explicit message saying "I don't promise." (See Figure 6 Appendix B).

[14] Participant 2's alternative to making a return promise was to send an explicit message saying "I don't promise." (See Figure 6 Appendix B).

are less than the payoffs they might realize in the subsequent modified dictator game if the Dictator cooperates, but the Recipient's outside option exceeds the payoff that he will receive if the Dictator decides not to cooperate with him. For example, opt out payoffs could be (7,7). See Figure 7 in the Appendix for an illustration of such a trust game. The trust game therefore gives the Dictator a clear reason to make a promise: if the Recipient takes her at her word, then the Recipient will opt in, which benefits the Dictator (see, e.g., Charness and Dufwenberg, 2006; Ederer and Stremitzer, 2009).

Since in our design subjects' promises took a conditional form–"I promise to cooperate if I am chosen to be the Dictator"–the promises that were made in our experiment were arguably more artificial than the promises that the Dictator would make in a trust game. But many promises in the real world have a conditional form. Insurance contracts, for example, always involve conditional promises. In our design, subjects effectively made promises to insure each other against the risk that each might end up in the vulnerable position of the Recipient. Most importantly, however, the trust game design was unsuitable for our purposes as it involves two reliance decisions: the Recipient's opt-in decision and his subsequent investment decision. Thus, a Recipient who opts in has relied on the promise by forgoing his outside option, even if he subsequently invests zero. And so, had we used the trust game, we wouldn't have been able to observe what happens if the Recipient decided not to rely on a promise. Furthermore, we would have only observed the effects of reliance on promising among recipients who opt in to the game. This subset of subjects may not be representative of the subject population, since recipients who opt in are likely to be more optimistic about the likelihood that the Dictator will cooperate than those that don't. In our setup, by contrast, all recipients made an investment decision. Finally, what matters is that subjects understand the messages that they send and receive to be promises and we have clear evidence that this is the case.[15]

3.2 Hypotheses

We are now in a position to formulate the hypotheses that flow from our theoretical model. First, we predict that behavior in the absence of a legal regime will depart significantly from the predictions of the classical model. More specifically, we predict 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. Since we predict that many dictators will make promises, it is optimal for dictators who made a promise to cooperate if they have high guilt parameters or, if recipients are willing to invest enough to induce them to cooperate, intermediate guilt parameters. Thus, H1.1 follows so long as there some dictators with sufficiently large guilt parameters (see Proposition 4). H1.2. follows so long as some recipients are confident that the dictator population consists of enough dictators with sufficiently large guilt parameters.

[15] As we explain in the results section, we also ran a Control treatment, which is identical to No Regime except that there is no communication stage so subjects cannot make promises to one another. A comparison of rates of cooperation in No Regime and Control shows that promising matters: 40% of dictators cooperate in No Regime compared to 20% of dictators in Control, and the difference is statistically significant (p=0.048, Wilcoxon rank-sum test). The difference is driven by dictators in No Regime who made a promise: 47% of such dictators cooperate compared to only 8% of dictators who made no promise.

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

Hypothesis 2 *Recipients exhibit a higher degree of confidence that dictators will cooperate as investment increases in No Regime (H2.1). Contingent cooperation rates will increase in No Regime as investment increases (H2.2).*

Explanation. Hypothesis 2 immediately follows from Proposition 6 as we predict that many dictators will make promises.[16]

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

Hypothesis 3 *The incidence and magnitude of overinvestment will be higher in No Regime than in Expectation Damages (H3.1). There will be less underinvestment and more efficient investment in Expectation Damages than in No Regime (H3.2).*

[16] In our example in Appendix A, 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.

Note that the "contingent cooperation rate" refers to the rate at which dictators are willing to cooperate contingent on their recipients choosing a particular investment level. The "actual cooperation rate" refers to the rate of cooperation given the investment levels actually chosen by their recipients. In other words, each dictator makes seven contingent cooperation decisions per round. One of those decisions–the decision corresponding to the investment level the recipient chose–is that dictator's actual cooperation decision.

Explanation. H3.1 follows from our model so long as some recipients hold beliefs about the dictator population that make overinvestment rational. That is, some recipients must believe that there is a high chance of facing a dictator with an intermediate guilt parameter that makes him willing to cooperate but only if investment is greater than one. H3.2 follows from our model so long as similar numbers of promises are made in Expectation Damages as in No Regime, and recipients aren't too confident that they face high-guilt dictators who will cooperate regardless of the investment level or recipients are sufficiently confident that they face intermediate-guilt dictators.

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. H4 follows from Propositions 8 and 7 so long as there are some low-guilt dictators who won't cooperate in the absence of legal enforcement regardless of the investment level, and a similar number of promises are made in Expectation Damages and 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 will follow if the above predictions are confirmed. This is because these predictions entail that investment decisions will be superior and rates of cooperation higher under Expectation Damages.

3.3 Procedure

We conducted 9 experimental sessions with a total of 140 student subjects. We used a between subject design, so subjects participated in only one treatment: 70 subjects participated in our Expectation Damages treatment and 70 subjects in our No Regime treatment.

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.[17]

Subjects were assigned to visually isolated computer terminals. Beside each terminal they found paper instructions, which are reproduced in 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).[18]

Each session consisted of two unpaid practice rounds followed by eight rounds,

[17] Treatments were evenly distributed across the two sites. At UC Santa Barbara, 30 subjects participated in the Expectation Damages treatment and 30 subjects participated in the No Regime treatment. At UC Berkeley, 40 subjects participated in the Expectation Damages treatment and 40 subjects participated in the No Regime treatment.

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

one of which was randomly chosen at the end of the experiment for payment based on the subjects' decisions in that round. Subjects were paid on average \$11.20 with earnings ranging between \$0 and \$15. In each round, subjects were anonymously matched with a randomly chosen participant. No participant interacted with the same participant more than once in any of the latter eight rounds. We achieved this by creating matching groups of exactly 10 subjects and having each subject play against the same participant during the two practice rounds. We also elicited recipients' beliefs about the likelihood their dictators would cooperate with them. In order to elicit beliefs in an incentive-compatible way, we paid recipients for their beliefs in a randomly selected round in accordance with Table 2 in Appendix B. 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. 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. Each round consisted of four steps.

Step 1. At the outset of the round, subjects were randomly and anonymously matched with another subject. Each subject in each pair then had the opportunity to exchange messages with one another sequentially in a randomly determined order. The participant who was selected to make the first communication decision (Participant 1) had to decide whether to send Message 1 to the other participant (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 choose not to send Message 1, she instead sent Participant 2 a message saying "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 chose not send Message 2, he instead sent Participant 1 a message saying "I don't promise." Notice that the conditional form of Message 1 means that Participant 1 only ends up promising to send Participant 2 money if Participant 2 makes a promise in return.[19]

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

Step 3. The Recipient then made his investment decision by selecting a level from zero to six. Prior to making this decision, he had to indicate his level of confidence

[19] Figure 6 in Appendix B shows which participants ended up making a promise depending on the messages sent during the communication stage. We might have instead required each subject to decide whether to make a promise to send money to the other if chosen to be Player A, regardless of whether the other sent him a return promise. Our reasons for making the promise of Participant 1 conditional on the receipt of a return promise were two-fold. First, since we are interested in the effects of reliance on promise keeping, we wanted to generate as many promises as possible, and the conditional form gives subjects stronger reasons to make promises than the unconditional form. Participant 1 has greater reason to make a promise, since his promise only becomes binding if he receives a promise in return. If Participant 2 receives a promise from Participant 1, she has greater reason to make a return promise in order to make Participant 1's promise binding. Second, we believe that the conditional form of the interaction better resembles the type of arrangement that subjects would arrive at if they had a free-form conversation. Had subjects been free to negotiate with one another, the likely result is that each would have only wanted to commit herself to being nice if she was assured of a return a commitment from the other. We expect that had we used unconditional promises, the less natural situation in which one participant makes a promise and the other does not would have arisen more frequently. Moreover, we find that there is no statistically significant difference in promise keeping between subjects who made conditional promises when in the role of Participant 1 and subjects who made unconditional promises in the role of Participant 2 ($p=0.69$, Wilcoxon rank-sum text).

Our design of the communication stage is similar to Vanberg's (2008). The main difference is that Vanberg used free-form messages, allowing subjects to exchange a sequence of instant messages. The advantage of computer-coded messages is that they more accurately capture promise-making without confounding it with other meaningful social interaction. The disadvantage of computer-coded messages is that they produce lower effect sizes (see, e.g., Charness and Dufwenberg, 2011).

that the Dictator would cooperate with him for each of the investment levels that he might have chosen. Table 2 in Appendix B shows how the Recipient's payoffs from guessing was determined for a given investment level as a function of his stated confidence level and the Dictator's actions.[20]

Step 4. After the Recipient made his investment decision, the Dictator then had to decide whether to cooperate by sending the Recipient money. Since we used the strategy method, the Dictator made this decision in ignorance of the Recipient's actual choice by indicating whether he 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. 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 beliefs.

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.

Given 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

[20]We use the belief elicitation mechanism used by Vanberg (2008).

subjects the opportunity to familiarize themselves with the game and the program interface before they played 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 and 57 subjects in Expectation Damages.

4 Results

The data 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 under No Regime and 456 decisions under Expectation Damages. However, the decisions made by subjects in each matching group consist of only one 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 happens when a Dictator has made a promise, our main results rely on aggregate data (though we sometimes also report differences focusing only on dictators who made a promise or recipients who received one). This is because the dictators who make promises in No Regime may differ systematically from those who make 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 could be driven by selection effects.

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 under No Regime, despite the fact that they lack a sufficient, 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 under No Regime, and this is significantly different from zero ($p=0.02$, Wilcoxon signed-rank test) (H 1.2).

A legal regime with expectation damages generates superior investment decisions. More interesting is our finding that 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 ($p<0.01$, Wilcoxon rank-sum test) (H3.2). There is also more efficient investment in Expectation Damages than in No Regime–58.8% compared to 21.8%–and the difference is statistically significant ($p<0.01$, Wilcoxon rank-sum test) (H3.2). But crucially, there is also less overinvestment in Expectation Damages than in No Regime–8.3% compared to 22.7%–and the difference is statistically significant ($p<0.05$, Wilcoxon rank-sum test). Moreover, average investment conditional on overinvesting is also lower under Expectation 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 (H3.1). Indeed, the average *increment* of overinvestment–a measure that combines both the incidence and magnitude of overinvestment[21]–is also considerably lower under Expectation Damages than under No Regime–0.12 compared to 0.56–and the difference is again statistically significant (p<0.01, Wilcoxon rank-sum test).

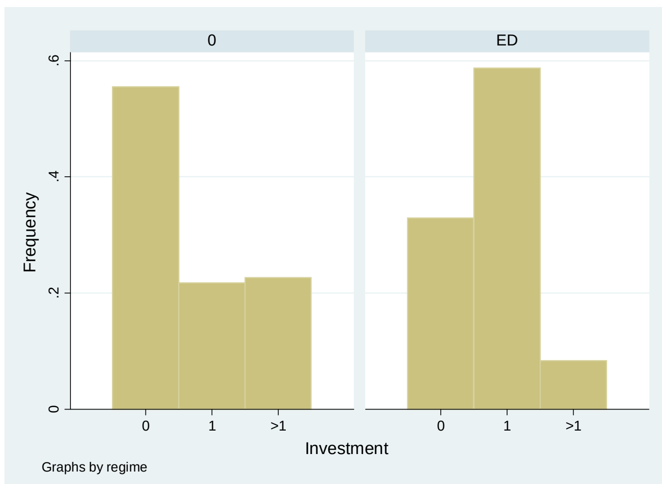

Figure 3: Incidence of Underinvestment, Efficient Investment, and Overinvestment.

These results are all consistent with our posited psychological lock-in effect. Some dictators are more willing to cooperate in No Regime the more their recipient has relied on a promise, and recipients anticipate this by overinvesting. In Expectation Damages, by contrast, there is no need to overinvest in order to increase a dictator's guilt from breaking a promise, because the legal regime ensures that a dictator has

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

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

	No Regime	ED	Z Stat
Efficient	0.22	0.59	$Z = 3.0^{***}$
	(0.29/216)	(0.49/228)	($p < 0.01$)
Underinvest	0.56	0.33	$Z = 2.6^{***}$
	(0.50/216)	(0.47/228)	($p < 0.01$)
Overinvest	0.23	0.08	$Z = 2.0^{**}$
	(0.42/216)	(0.28/228)	($p < 0.05$)

* Pooled data from all sessions and all rounds. Standard error 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.

a sufficient self-interested reason to keep her promise, so long as the recipient invests at least one, the first-best level.

Overinvestment under No Regime appears to be driven by a psychological lock-in effect. If this psychological lock-in effect is really driving our results, then we should expect to see that dictators do in fact become more willing to cooperate and recipients become more confident that dictators will cooperate when recipients overinvest as Hypothesis 2 predicts. We have data to test this hypothesis because we employed the strategy method to elicit the dictators' preferences for all possible investment levels, and we elicited recipients' beliefs about the likelihood that a dictator will cooperate for all possible investment levels. Our theoretical model 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 psychological lock-in effect depends on the existence of this intermediate group of dictators whose preferences depend on their investment level.

More precisely, if we suppose that 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, as higher investment levels induce dictators with lower guilt parameters to keep their promises while having no effect on the willingness of those with higher guilt parameters who were already willing to keep their promises at the lower level of investment.[22] Thus, to test Hypothesis 2 we compare Recipients' average confidence in dictators' cooperation in No Regime for investment of zero with recipients' average confidence in dictators' cooperation for investment of 6. Confidence increases from 1.8 to 2.2 where we measured confidence on a scale from 0 to 4 (see No Regime graph in Figure 4). This difference is statistically significant (p=0.03, Wilcoxon sign-rank test) (H2.1). We also observe that average contingent cooperation rates increase from 41% for investment of 0 to 49% for investment of 6 (see No Regime graph in Figure 5). 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) (H2.2).[23]

[22]Note that in Figure 5, contrary to this prediction, we also 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 (Wilcoxon sign-rank test, p=0.13, two-sided t-test, p=0.14). 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.

[23]We report the one-sided t-test where we have a directed hypothesis and the two-sided t-test where we don't.

Note that the same subjects act as dictators and recipients. So 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 under Expectation Damages than No Regime: 62.3% compared to 40.3%. The difference is statistically significant (p=0.02, Wilcoxon rank-sum test) (Hypothesis 4). This is not particularly surprising. So long as dictators make promises, thus ensuring that there is legal enforcement of relied-upon promises, this result is consistent with the predictions of the standard self-interested rational-actor model as well as our model in which dictators experience guilt from breaking their promises. Indeed, the differences are more stark when we focus only on dictators who made a promise: 93.2% cooperate under Expectation Damages, while 47.5% cooperate under No Regime. Again, the difference is statistically significant ($p<0.01$, 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.

These results, however, mask one difference in contingent cooperation rates that is not predicted by the standard model. In the event the legal regime is not triggered because a recipient invests zero, contingent cooperation rates are considerably lower under Expectation Damages than under No Regime: 18.9% compared to 40.7% in aggregate, and 24.7% compared to 46.9% among 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.) This suggests that the presence of a legal regime *crowds out* voluntary promise-keeping.[24] In the terms of the model, the legal regime seems

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

to reduce the guilt of dictators, holding payoffs constant (i.e., it seems to reduce the guilt parameter γ). Recipients, moreover, anticipate this effect. Average recipient confidence for zero investment is higher under No Regime than Expectation Damages: 1.8 compared to 1.1 in aggregate, and 2.1 compared to 1.4 among dictators who made a promise. Both differences are statistically significant (p=0.03 for both, Wilcoxon rank-sum test).

This crowding out seems unlikely to have a large effect on overall cooperation rates, however, since recipients can avoid the effect, 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 underinvest if they received a promise (compared to 52.2% in No Regime and 89.6% in Expectation Damages if there was no promise)

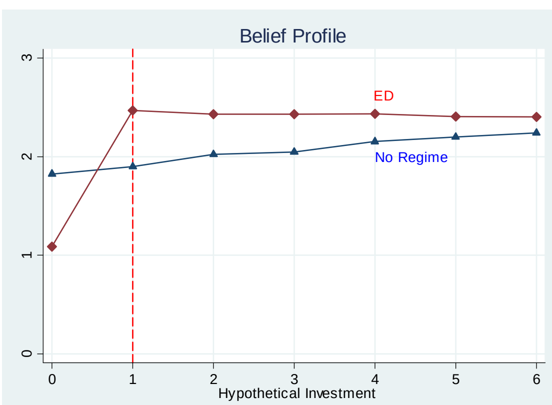

Figure 4

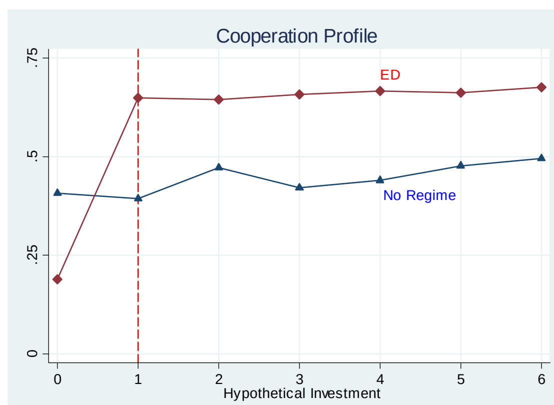

Figure 5

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 \$22.9 under Expectation Damages compared to \$21.7 under No Regime and the difference is significant (p=0.01, Wilcoxon rank-sum test). The difference is even starker when we look only at subject pairs where the dictator made a promise: \$23.9 compared to \$21.9 (p<0.01, Wilcoxon rank-sum test). (There is no statistically significant difference in average joint payoffs among subject pairs 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 in subject pairs from \$5.90 to \$3.15, a statistically significant reduction (p=0.02, Wilcoxon rank-sum test). The difference is even more marked when we focus on subject pairs in which the dictator made a promise: \$5.3 compared to \$0.1 (p<0.01, Wilcoxon rank-sum test). (Again, there is no statistically significant difference in average joint payoffs among subject pairs in which the dictator didn't make a promise.) Average payoff differences are a measure of risk resulting from opportunism. Before subjects are assigned their roles, they don't know whether they will be in the position of the dictator or recipient. That is, they don't know whether they will be in a position to exploit the other subject or instead be exploited. Our results suggest that at this ex ante stage, a risk-averse subject would prefer Expectation Damages to No Regime, since it reduces the risk that he will be exploited by the other ex post. (H5.2)

Competing Hypotheses. Our model assumes that the Dictator is responsive to the Recipient's investment level only if she made a promise. Formally, if $p = 0$, the Dictator cares only about her material payoff. Likewise, our model predicts that the Recipient will overinvest to make the Dictator more likely to cooperate only if the

Dictator made a promise. However, there is plenty of evidence in the experimental literature on cooperative behavior in dictator and ultimatum games as well as in public goods games that, even in contexts where no promises are made, cooperation exceeds zero because of efficiency or equity preferences. And if the Dictator exhibits such preferences, her cooperation may increase as investment increases, and the Recipient may accordingly overinvest to make the Dictator more willing to cooperate, even in the absence of a promise.

Efficiency Preferences. By equations (7) and (8), the difference in players' joint payoffs when the Dictator cooperates, $a = 1$, and when she doesn't cooperate, $a = 0$, is $3 + \frac{5}{4}$ for $i \leq 1$ and $3.5 + \frac{3}{4}i$ for $i > 1$. Thus, the difference in social surplus is higher when the Dictator cooperates, $a = 1$, than when she doesn't cooperate, 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 (7) and (8), 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, $a = 1$, than when she doesn't cooperate, 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 2 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 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.

One reason to believe that our model likely captures an important mechanism behind the observed effect can be found in the prior literature on the effect of promisees' expectations on promise-keeping. While promisees' expectations per se do not seem to matter (Vanberg, 2008; Ellingsen et al, 2010), they do matter to a promisor who induced those expectations in the promisee by making a promise (Ederer and Stremitzer, 2015). To the extent that we should expect a promisee's reliance on a promise to have 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 prior findings.[25]

The primary focus of the present study is to explore the possibility that enforcement could reduce overinvestment as well as underinvestment, rather than to identify

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

the psychological mechanism behind this effect. However, we also ran an additional Control treatment that was identical to the No Regime treatment with the exception that participants were not able to send messages to each other. This allows us to obtain direct empirical support for the stronger claim that there is something distinctively *promissory* about our lock-in effect–that is, that dictators' sensitivity to recipients' investment choices is triggered, or, at least, intensified when they made a promise. Consistent with this theory 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 more detailed discussion that makes use of the belief and contingent cooperation data see Appendix C.

5 Discussion

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 promisor more likely to keep it. Promisees, moreover, seem to anticipate this effect and so *strategically* rely on promises in order to make their promisors more likely to keep their promises. As a result, enforcing relied-upon promises with expectation damages has the unexpected effect of reducing overinvestment, as well as reducing underinvestment and increasing the incidence of efficient investment and increasing rates of cooperation. Hence, crowding out the extra-legal mechanism of psychological lock-in by the legal regime has a positive effect in our setting. This result cannot be explained by assuming that subjects have standard, self-interested preferences. But it can be explained by our model, according to which promisors are more motivated to keep their promises the more their promisees have relied on them.

It is notable that Expectation Damages generates superior investment and cooperation decisions than those generated by No Regime, even though Expectation Damages causes a reduction in the numbers of promises that get made: 82% of subjects make promises under No Regime compared to 64% under Expectation Damages. The difference is significant only at 10% however (p=0.06, Wilcoxon rank-sum test). This reduction in promise making is unsurprising given that promises are cheap-talk under No Regime but not under Expectation Damages. But the fact that we observe this reduction doesn't undermine any of our results: we would expect Expectation Damages to do even better if promise rates were the same across the two regimes.

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. We observe that contingent cooperation under Expectation Damages rises sharply at an investment level of one when a promise was made and remains high for higher investment levels. We also observe 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. This suggests that subjects understood their payoffs under Expectation Damages. Under No Regime contingent cooperation and recipients' average confidence levels rise more steadily as investment increases consistent with our posited psychological lock-in effect.

Finally, a possible limitation of our results is that we obtained our data using the strategy method, which means that we elicited dictators' preferences for all possible investment levels, including those not actually chosen by the recipient. We could have instead simply asked dictators to make their selection after observing the investment level that the recipient in fact chose. Though employing this "direct-response" method would have generated considerably less data, it more closely resembles real-world choice situations, and it is natural to wonder whether it would have generated the same results. While this is an important question, we don't expect that we would have obtained qualitatively different results had we employed the direct-response method. Based on a survey of experimental studies that employ both methods, Brandts and Charness (2011) find that "there are significantly more studies that find no difference across elicitation methods than studies that find a difference" (p. 387). Indeed, it is plausible to suppose that the strategy method worked against us. Any guilt a dictator experiences when she makes a decision to break a promise to cooperate may be intensified when she knows for sure that the result of that decision will certainly materialize. If so, then using the "direct-response" would exacerbate the effects of reliance on dictators' willingness to keep their promises. As Brandts and Charness (2011) note, there is some evidence that emotions run higher when decisions are elicited under the direct-response method.

6 Conclusion

Our results suggest 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 their promisors. Reliance on promises matters, because many promisors are more likely to keep promises that have been relied upon. And promisees anticipate that reliance has this effect, and accordingly some promisees overinvest in order to prod their promisors into keeping their promises.

How plausible is it that we will observe this psychological lock-in effect outside the laboratory? It is plausible to suppose that when individuals make promises to one another, their reliance decisions will exhibit psychological lock-in. But what about promises made by sophisticated, collective entities like corporations? It is perhaps less plausible to suppose that such actors exhibit the guilt-averse preferences that would give a promisee a strategic incentive to over-rely on a promise. But our results tell us something about people's moral intuitions about promising: people seem to think it is worse to break a promise the more it has been relied upon. If that is the case, it is plausible to suppose that when a corporation is seen to break a promise, its reputation will suffer more the more the promisee relied on it. And this would give a promisee a strategic reason to over-rely on a promise made by a corporation, even if the corporation's preferences are narrowly self-interested.

References

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

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

——— (2009): "Dynamic Psychological Games," *Journal of Economic Theory*, 144(1), 1–35.

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

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

Braver, S. L. (1995): "Social Contracts and the Provision of Public Goods," in *Social Dilemmas: Perspectives on Individuals and Groups*, ed. by D. Schroeder. Praeger, New York.

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

Charness, G.;, M. ((2010)): "Bare Promises: An Experiment.," *Economics Letters*, 107, 281–83.

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

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

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

Cooter, R., and M. A. Eisenberg (1985): "Damages for Breach of Contract," *California Law Review*, 73 (5), 1432–1481.

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

Ederer, F., and A. Stremitzer (2015): "Promises and Expectations," *Cowles Foundation Discussion Papers No 1931.*

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

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

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

Ellingsen, T., M. Johannesson, S. Tjøtta, and G. Torsvik (2010): "Testing Guilt Aversion," *Games and Economic Behavior*, 68(1), 95–107.

Farber, D., and J. Matheson (1985): "Beyond Promissory Estoppel: Contract Law and the "Invisible Handshake."," *52 U. Chi. L. Rev.*, p. 703.

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

Frey, B., and R. Jegen (2001): "Motivation Crowding Theory," *Journal of Economic Surveys*, 15(5), 589–611.

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

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

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

Hoppe, E. I., and P. W. Schmitz (2011): "Can Contracts Solve the Hold-Up Problem? Experimental Evidence," *Games and Economic Behavior*, 73(1), 186–199.

Ismayilov, H., and J. J. M. Potters (2012): "Promises as Commitments," *Tilburg University, Center for Economic Research, Discussion Paper*, 2012-064.

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

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

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

Ostrom, E., J. Walker;, and R. Gardner (1992): "Covenants With and Without a Sword: Self-Governance Is Possible," *American Political Science Review*, 86(2), 404–417.

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

Sally, S. (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.

Selten, R. (1967): "Die Strategiemethode zur Erforschung des eingeschraenkt rationalen Verhaltens im Rahmen eines Oligopolexperiments," in *Beitraege zur experimentellen Wirtschaftsforschung*, ed. by H. Sauermann., pp. 136–168. Tuebingen: Mohr.

Shavell, S. (1980): "Damage Measures for Breach of Contract," *Bell Journal of Economics*, 11, 466–490.

——— (1984): "The Design of Contracts and Remedies for Breach," *Quarterly Journal of Economics*, 94, 121–148.

Sloof, R. e. a. (2003): "An Experimental Comparison of Reliance Levels under Alternative Breach Remedies," *RAND Journal of Economics*, 26(3), 205–222.

——— (2006): "Breach remedies, reliance and renegotiation," *International Review of Law and Economics*, 34), 263–296.

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

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

Wilkinson-Ryan, T., and D. A. Hoffman (2015): "The Common Sense of Contract Formation," *Stanford Law Review*, 67, 1269–1301.

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

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

Yorio, S., and S. Thel (1985): "Promissory Basis of Section 90," *101 Yale L. J.*, p. 111.

APPENDIX A

Proof of Proposition 4 If there was no promise, the equilibrium follows immediately from inserting $p = 0$ into the the utility function (9) 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 (6) and (10) imply that $g(\cdot)$ is increasing in i. Thus, $\gamma^c(i)$ is decreasing in i. If $\gamma < \gamma^c(\bar{i})$, then no investment level in the Recipient's choice set will induce the Dictator to cooperate, and therefore $\left(\widehat{a}, \widehat{i}\right) = (0, 0)$. But if $\gamma \geq \gamma^c(\bar{i})$, then there exists a critical investment level $i^c \leq \bar{i}$ such that whenever $i \geq i^c$, a Dictator for whom $\gamma \geq \gamma^c(i^c) \geq \gamma^c(\bar{i})$ will cooperate. Thus, $\gamma \geq \gamma^c(e)$ 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 (2), the Recipient will always choose e, and the Dictator will cooperate: $\left(\widehat{a}, \widehat{i}\right) = (1, e)$. If, by contrast, $\gamma^c(\bar{i}) \leq \gamma < \gamma^c(e)$ then $i^c \in (e, \bar{i}]$. As $\max_{i \in [i^c, \bar{i}]} \pi_R(1, i) = \pi_R(1, i^c)$, the Recipient will choose $\widehat{i} = i^c$ and so $\left(\widehat{a}, \widehat{i}\right) = (1, i^c)$ if $\pi_R(1, i^c) > \pi_R(0, 0)$, and $(\widehat{a}, \widehat{i}) = (0, 0)$, otherwise.

Simple example exploring the basic tradeoffs that result from introducing uncertainty over Dictator types.

No Legal Enforcement. Suppose that material payoffs are given by our particular specification in (7) and (8), and that the Dictator's guilt function is linear so that her utility in the event that she made a promise is given by:

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

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 $\gamma = \frac{3}{8}$. It can therefore be seen that 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, $p_3 < 4(p_1 + p_2)$, *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 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, $p_3 \geq 4(p_1 + p_2)$, *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 invests one,* $\widehat{i} = 1$, *so* $\left(\widehat{a}^m, \widehat{i}\right) = (0, 1)$ *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:

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

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: $\widehat{i} = 2$ if

$$U_R(2) \geq U_R(0) \Leftrightarrow p_1 \leq 3p_2$$

and $\widehat{i} = 0$ 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 $\widehat{i} = 1$ otherwise. $\widehat{i} = 1$ otherwise. ■

When the Recipient is sufficiently confident that the Dictator is an intermediate type, holding constant the likelihood that he is a high-guilt type, then he invests 2 in order to induce the intermediate type to cooperate. But his 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 there is instead a high chance he faces a low-guilt type, and a sufficiently low chance that he faces an intermediate-guilt type, then it likely the Dictator won't cooperate regardless of his investment level, making it optimal for him to invest zero.

Legal enforcement. When a legal regime with expectation damages is introduced, Proposition 5 implies that the Recipient will always invest 1, since expectation damages ensure that the Recipient always receives $\pi_R(1, i)$ so long as he invests at least one no matter what the Dictator does, and the Dictator will always prefer to cooperate: $\left(\widehat{a}_L, \widehat{i}_L\right) = (1, 1)$.

In the presence of a legal regime with expectation damages, all three types of Dictator cooperate so long as the Recipient invests one. Thus, the uncertainty about the Dictator's type makes no difference to the equilibrium when there is a legal regime with Expectation Damages. Proposition 8 follows immediately 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,* $\left(\widehat{a}_L, \widehat{i}_L\right) = (1, 1)$.

APPENDIX B: FIGURES & TABLES

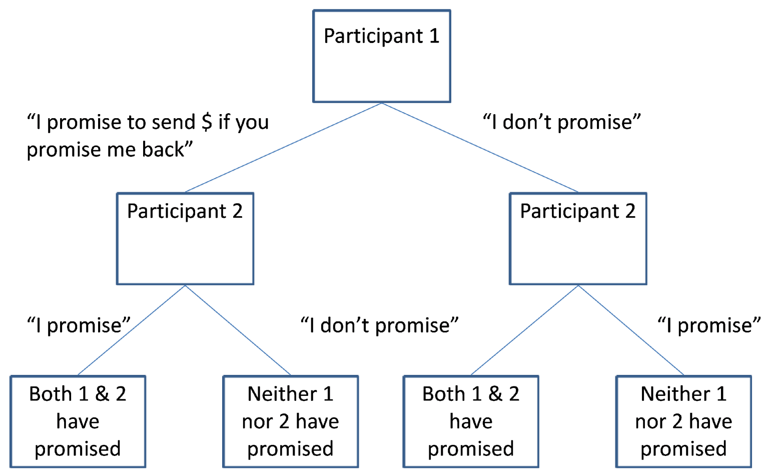

Figure 6: Design of the Comunication Stage.

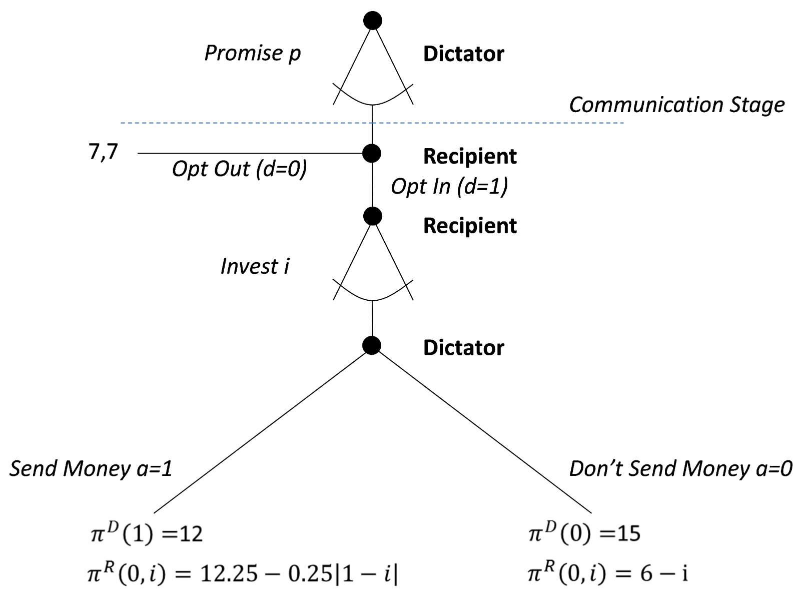

Figure 7: Alternative Promising Technology: Trust Game

Player A Will...	Certainly Choose *Send Money*	Probably Choose *Send Money*	Unsure	Probably Choose *Don't Send Money*	Certainly Choose *Don't Send Money*
Your earnings if the other player chooses to *send you money*	\$0.65	\$0.60	\$0.50	\$0.35	\$0.15
Your earnings if the other player chooses to *not to send you money*	\$0.65	\$0.60	\$0.50	\$0.35	\$0.15

Table 2: Player B's Stated Beliefs About the Likelihood that Player A will Send Money

APPENDIX C: Alternative Hypotheses

Running 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, allows us to obtain 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 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, 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 could have an alternative explanation for our psychological lock-in effect.[26]

[26] By supposing that the three component's of the Dictator's utility function are additive, we

Thus, the difference in the Dictator's utility when she cooperates compared to 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, a promise shifts 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

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. Rather, *both* have to be present to generate psychological lock-in. 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.

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 sign-rank test, p=0.04), and the graph of recipient confidence is everywhere non-decreasing (Figure 8).

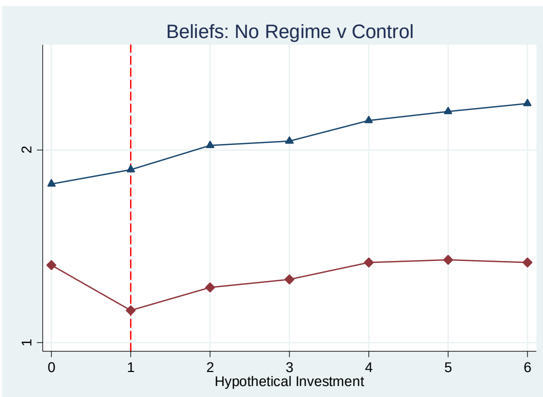

Figure 8

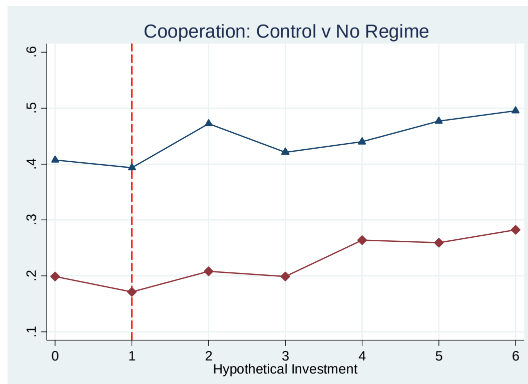

Figure 9

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 9), 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.

APPENDIX D: INSTRUCTIONS*[27]

[27]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 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.

<u>Payoffs if Player A **did not promise** to send Player B money</u>

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

<u>Payoffs if Player A **promised** to send Player B money</u>

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	