Renegotiation Behavior and Promise-keeping Norms

By Erin Krupka, Stephen Leider and Ming Jiang*

Contracts rely on informal agreements and renegotiation, which both rely on norms. We test how promises and renegotiation impacts norms in a trust-based game using an experiment. We present a theoretical framework where successful (rejected) renegotiation strengthens (weakens) the norm. Our results support the framework's predictions: most subjects make promises, and promises are largely fulfilled even when costly. The mere opportunity to renegotiate has a causal effect on trustworthiness and successful renegotiation increases trustworthiness. As predicted, many subjects do not renegotiate even though there is no strategic downside. Heterogeneity in beliefs in the norm strength predicts which subjects renegotiate.

JEL: C9; D23

Keywords: Renegotiation; Norms; Promises

I. Introduction

Not all contingencies can be anticipated or specified, so formal contracts often rely on complementary informal agreements and renegotiation. Both of the latter rely on the force of norms - a shared understanding of how each party ought to behave.[1] "Many transactions will potentially be too costly to undertake if the

* Krupka: School of Information, University of Michigan, ekrupka@umich.edu. Leider: Ross School of Business, University of Michigan, leider@umich.edu. Jiang: Antai College of Economics and Management, Shanghai Jiao Tong University, mjiang@sjtu.edu.cn. We thank Emma Enache, Andrew Barton and Jennifer Lyu for outstanding research assistance. In addition, we wish to thank Martin Dufwenberg and George Akerlof, participants of the ERINN conference held at Oxford and LSE in 2019 as well as participants of the Economic Science Association conference 2016. This research has been approved by the University of Michigan IRB.

[1]For example, informal promises can have a large effect on behavior (Charness and Grosskopf, 2004; Fehr, Hart and Zehnder, 2011; Feltovich and Swierzbinski, 2011; Miettinen, 2013), even when fulfilling the promise entails a personal cost (Ellingsen and Johannesson, 2004; Charness and Dufwenberg, 2006; Vanberg, 2008). Recent work shows that norms of promise-keeping underpin these informal agreements

participants cannot rely on efficient and equitable adaptation to those unforeseen contingencies. Note that such reliance will necessarily involve blind faith....both sides ... must *ex ante* have some idea of the meaning of appropriate or equitable fulfillment of the contract." (Kreps, 1990)[2] Equitable adaptation, and the *ex* ante 'ideas' which underpin those adaptations, are articulated in the form of moral principles that have wide applicability. Social norms reflect the collective understanding of the dominant principle for the interaction of interest. Though previous work has documented that norms influence behavior in settings with informal agreements, we pose the question of how promises and renegotiation impact the norms governing the interaction.

Behavioral accounts for renegotiation focus on the psychology of "contracts act as reference points" and the accompanying norms of fairness (Hart, 1995; Charness and Grosskopf, 2004; Feltovich and Swierzbinski, 2011; Miettinen, 2013; Vanberg, 2008; Charness and Dufwenberg, 2006).[3] Hart and Moore (2008) model the initial agreement as setting a reference point for what determines fair treatment. If one party uses discretion or renegotiation to give the other party a worse outcome than he expected under the initial agreement, then this party may feel aggrieved and use ex-post actions to punish the other. They consider several ways renegotiation interacts with feelings of aggrievement. Each party could feel entitled to all of the newly generated surplus, or could require that the renegotiation make both parties better off. Hart and Moore (2008) emphasize that the flexibility allowed by discretion and renegotiation is in tension with the costs of potential aggrievement.

Halonen-Akatwijuka and Hart (2013) more explicitly include renegotiation into

(Kessler and Leider, 2012; Krupka, Leider and Jiang, 2016).

[2]Macaulay (1963) documents the under-specification of many manufacturing contracts; similarly, Carlton (1986) suggests that for many industrial transactions the "contracts specify neither price nor quantity". See also Kessler and Leider (2012) and Lyons (1996). Employment contracts often specify only hours, duration, and compensation. Service contracts are often similarly simple (e.g. hourly rate or fixed price contracts) and, generally, neither specify particular behaviors nor make provisions contingent on potentially verifiable information (see Eggleston, Posner and Zeckhauser (2000)).

[3]Renegotiation-proof contracts, when possible, typically require substantial distortions and inefficiency (Hart and Moore, 1988; Dewatripont, 1988).

the Hart and Moore framework, showing that the two sides will have divergent beliefs about how much of the surplus they should receive from renegotiation which can lead to surplus-destroying "shading" of performance. These concerns can then significantly distort the initial contract. Herweg and Schmidt (2014) similarly examine how the original contract creates a reference point for interpreting changes as gains or losses. Loss aversion creates a wedge between what the buyer and seller will accept that results in materially inefficient renegotiation or no renegotiation at all.[4]

In these models the primary barrier to renegotiation is a strategic concern about how parties will react to deviations from a reference point. Specifically, parties may choose not to ask for renegotiation if they fear the other party will react to perceived violations of fairness norms created by the renegotiation. However, these models have less explanatory power in settings where renegotiation is not primarily driven by strategic concerns.

The empirical evidence for the "contracts as reference points" framework shows that initial contracts matter for determining behavior. However, they also struggle to fully explain important and systematic patterns in the data. Fehr, Hart and Zehnder (2015) introduce informal agreements and renegotiation to the "contracts as reference points" experiments that compare rigid and flexible contracts (Fehr, Zehnder and Hart, 2009; Fehr, Hart and Zehnder, 2011).[5] They show that informal agreements reduce "shading", but do not eliminate the trade-off between contractual rigidity and flexibility. Renegotiation allows rigid contracts to perform better - as mutually beneficial changes in the high cost state can lead to more transactions, while avoiding the aggrievement that flexible contracts

[4]Loss aversion captures a different psychological notion than that of reference points. In this case there is an asymmetry between how we experience losses vs gains. People tend to prefer avoiding losses to acquiring equivalent gains (Tversky and Kahneman, 1979).

[5]In this setting, there is *ex ante* uncertainty over whether the seller's cost is high or low, and parties can sign either a rigid fixed price contract or a flexible price range contract (with final price decided by B after realized costs). Flexible contracts allow trading in both states of the world, but create an ambiguous reference point that can create aggrievement. In two treatments the buyer can make an informal agreement stating what price they will set for each cost realization and the buyer can unilaterally replace the existing contract rather than renegotiate.

generate in the low cost state. Additionally, opportunistic renegotiation leads to retribution. However, the opportunity to renegotiate also appears to influence the way that the flexible contract creates a reference point and aggrievement even when renegotiation is not taken advantage of.

Bartling and Schmidt (2015) also find evidence for contracts to create reference points in a setting with renegotiation. In their experiment an initial contract specifies a price and a design, however *ex post* another design may be more valuable - requiring renegotiation. The initial contract prompts suppliers to make less aggressive renegotiation offers. Bartling and Schmidt (2015) conclude that multiple behavioral forces beyond the reference point effect are needed to explain their data and that internalized norms may explain their observations.[6] Iyer and Schoar (2015) provide evidence for the importance of social norms from two field experiments on renegotiation. Their evidence suggests that norms can also prevent renegotiation. In the field experiments, both tailors asked to handle a rush order and vendors asked to cut the price for a custom order often declined to renegotiate when doing so would require a transaction with terms deemed to be "inappropriate".

The existing literature offers some explanations for the role of informal agreements to impact when renegotiation is likely to take place and what gets renegotiated. This body of work suggest that norms play a central role in supporting the informal and formal aspects of contracting and renegotiation. However, across papers, the authors note that multiple behavioral forces must be combined in order to shed light on the "when", "what" and "why" of contracting and renegotiation.[7]

In this paper, we begin with the "behavioral" premise that contracting is a social interaction and, as such, is governed by the *grammar* of social interaction - i.e by social norms (Bicchieri, 2005). Norms represent shared understanding of morally

[6]For example, in many cases suppliers are willing to make the change without renegotiation, but only when doing so lowers their costs. They argue that this reflects internalized social norms regarding when renegotiation is justifiable (eg. when costs are higher).

[7]Fehr, Hart and Zehnder (2015) write, "Most important, the experimental setups ignore real-life aspects of trading relationships such as informal agreements and renegotiation."

or socially acceptable behavior and are often described with a key moral principle: e.g. fairness, trustworthiness, reciprocity, promise-keeping, etc.(Charness and Dufwenberg, 2006; Krupka, Leider and Jiang, 2016). In a setting with uncertainty, we ask whether we can use a norms framework to explain the outcomes we observe both when renegotiation is and is not present. Our approach is to argue that an (un)successful renegotiation increases (decreases) the perceive importance of following the dominant moral principle.[8]

We proceed by exploring how promises and renegotiation impacts the importance of norms of trustworthiness in a trust-based game. Our norms model characterizes how making promises and opportunities for renegotiation affect the appropriateness of taking specific actions in our setting. The model captures the intuition that decision makers derive utility from complying with norms, and that the norms may be strengthened or weakened by making promises and when renegotiation is successful or unsuccessful. We show that such a model for behavior is both flexible and powerful – making clear predictions based on easily observable features of the norm.

Using an experiment, we show that many subjects' behavior can be explained as acting to "manage" the norms (and hence increase utility for norm compliance and decrease the disutility for non-compliance) rather than simply maximizing payoff or managing strategic considerations. Finally, we can explain a range of behaviors with only assuming a normative component to the final act of (non-) trustworthiness.

To test these predictions, we modify the trust game presented in Charness and Dufwenberg (2006) to introduce informal agreements and renegotiation in a way that eliminates strategic concerns for the second player. We modify the first stage so that Player A ("she") and B ("he") are making a simultaneous decision. A is able to choose whether she wishes to be In or Out for the game,

[8]Note that we do not assume that there is a separate norm directly governing the renegotiation itself. Rather the impact of the renegotiation is channeled through its impact on the primary norm.

i.e. if A is willing to trust B. At the same time, B is deciding whether to make a non-binding promise to Roll. If A chooses In, then the payoffs are determined by whether B chooses to be trustworthy and roll the die. Because these initial decisions are made simultaneously, making a promise has no strategic value to B, since it cannot affect A's decision. We call this the risky trust game.

The Renegotiation treatment adds one more modification to the game. After receiving a negative shock (which makes it more costly to roll), B is given a surprise opportunity to ask for renegotiate of the payoffs so that A takes on half the extra cost (but only if B follows through and rolls). If A refuses to renegotiate, the game continues with the payoffs unmodified. Since A has no opportunity to punish B, and refusing to renegotiate creates the same material payoffs as obtained if B did not ask for renegotiation, there is no strategic reason for B to forgo requesting to renegotiate.

In order to test the impact of norms on behavior in these two settings (Baseline and Renegotiation), we use an incentivized norm elicitation method (Krupka and Weber, 2013). This creates an empirical proxy for how the norm of trustworthiness (via the decision to roll a payoff determining die) is altered in the presence of promises and/or (un)successful renegotiations. In addition, we connect an individual's play in the risky trust game to their particular beliefs about the norm under different paths of play.

We find broad support for a behavioral model in which norms figure prominently. Despite the lack of strategic benefit, more than 80% of subjects make a promise to be trustworthy *and* promises are associated with greater trustworthiness. Subjects who promised are more than 40% likely to be trustworthy when the costs to keep the promise are the greatest. In our framework this is reflected by subjects who intend to be trustworthy and make a promise, thereby strengthening the social norm, in order to receive a greater utility for being trustworthy (and fulfilling the norm).

Second, the opportunity to renegotiate has a causal effect on trustworthiness

and the outcome of renegotiation matters - subjects are nearly 35% more likely to be trustworthy after a successful renegotiation. Surprisingly, only about 50% of subjects ask for renegotiation. Our norms data explains these observations. On average there is a stronger norm to be trustworthy (and roll) after an accepted renegotiation, compared to a rejection. This is consistent with our foundational assumption that renegotiation outcomes influence behavior through increasing or decreasing the perceived strength of the norm (reflecting the primary moral principle of promise-keeping).

However, while there is overall agreement about the general impact of renegotiation outcomes on the norm, there is substantial heterogeneity in beliefs about the strength of the norm after a rejection and this heterogeneity helps explain individual behavior. The most likely subjects to ask for renegotiation are those who think there is a strong norm to roll only for an accepted renegotiation, while those least likely to *ask* for renegotiation are those that think there is a strong norm in all cases (i.e. even after a rejection in the renegotiation). Given their beliefs, the failure to ask for renegotiation can be explained by trying to manage the relevant norm: they are likely to be trustworthy in any case, so they would prefer not to ask for a renegotiation than to be rejected and receive less utility for being trustworthy under a weaker norm.

Our contribution is to apply a theoretical framework where social norms offer an explanation for important patterns of behavior in trustworthiness and renegotiation. We can use this framework to explain why subjects make promises, why promises are associated with being more trustworthy, how renegotiation outcomes influence trustworthiness, and how heterogeneity in beliefs about the norms predict (non-)renegotiation. In doing so, we also introduce an empirical strategy to construct a proxy for norm strength, allowing for individual-level predictions.

These results have broad implications regarding how renegotiation affects norms and trustworthiness, and provide important refinements to our knowledge of how to structure contracts. Subjects take actions that "manage" relevant norms and,

in doing so, provide compelling evidence for modeling contracting with a social norms framework. We push further the view reflected in the leading quote by Kreps (1990) and in the vast majority of empirical work on norms and contracts: A vision of norms as rules of conduct that are established outside of the particulars of the transaction or how the transaction unfolds.[9] We extend this view by characterizing norms whose strength is conditional with respect to the unforeseen contingencies and behaviors of the interacting parties. Similarly, we follow the literature in treating norms as one common understanding, but we extend the value of this approach and use meaningful individual heterogeneity in *beliefs* about norm strength to predict breakdowns in the interaction. As a consequence, we expand the impact that norms likely have in these settings and are able to account for a wide suite of behaviors. The foundational assumption that renegotiation influences social norms by increasing or decreasing the strength of the dominant moral principle captured by the norm is applicable to any transaction (or game) where moral principles play a role.

II. The Setting

This section describes the environment our subjects encountered and sets the stage for our theoretical framework and predictions. We modify the trust game used in Charness and Dufwenberg (2006). In Figure 1 we describe the risky trust game so that it maps to the experimental design. The game proceeds in four stages (with an additional stage added in the Renegotiation treatment). At the start of the game, both players are aware of the structure of the game and the distribution of possible X values (except for the surprise opportunity to renegotiate, i.e. the presence of Stage 2.5, in the Renegotiation treatment).

Stage 1 (Trust and Promise): In the first stage, Player A chooses whether

[9] There is good reason to begin with such assumptions. The impartial legitimacy of a norm established outside of the specific interactions greases the wheels of trust; establishes reference points to guide behavior; can be used by parties to create reputation, assess ex ante intentions, or justify the legitimacy for renegotiation.

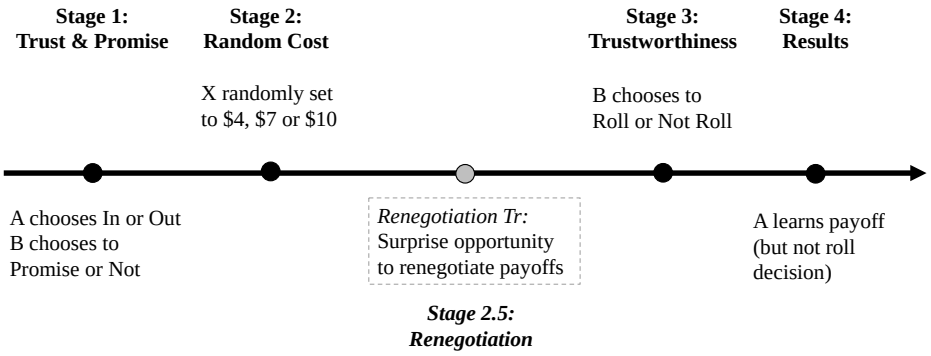


FIGURE 1. GAME TIMELINE WITH/OUT RENEGOTIATION.

to trust Player B by choosing In to continue the game, or not to trust B by choosing Out to end the game. If A chooses Out the game ends and both A and B receive $2. If A chooses In, the game proceeds to Stages 2 through 4. Simultaneously, B is choosing whether to make a non-binding promise to roll a virtual die to (potentially) generate a positive payoff for A in Stage 3. Rolling the die will be costly for B relative to not rolling and giving A a payoff of zero. B's promise cannot influence A's decision to trust, and hence the promise has no strategic value because it is made at the same time that A chooses In/Out. We argue that the promise changes the relevant norm that governs B's decision to be trustworthy in Stage 3. At Stage 1 both players know the distribution of potential costs, but the value is set randomly in Stage 2.

Stage 2 (Random Cost): In Stage 2, the random cost for Player B to be trustworthy is determined. The variable X is set to one of three values from the known distribution: with a 40% chance $X = \$4$, 40% chance $X = \$7$, and 20% chance $X = \$10$. Both players are informed of the realized value of X. X represents Player B's payoff if they choose to be trustworthy, and hence lower values of X make trustworthiness more costly.[10] By probabilistically varying the payoff associated with rolling the die, we change the cost of rolling and honoring

[10]We can think of low values of X as a bad draw from nature (eg. market forces that cause the value of your mortgaged home to decline) and thus, increasing the costs to keeping one's promise.

the promise and examine the impact of those costs on fulfilling the agreement. When $X = \$10$ there is no cost to be trustworthy, while when $X = \$4$ there is a substantial cost.

Stage 3 (Trustworthiness): In Stage 3, Player B observes the value of X, and chooses whether to be trustworthy and Roll the virtual die. If B chooses not to roll then Player A will receive \$0 and B will receive \$10. If B chooses Roll they will receive \$X for sure. Player A's payoff, however, is determined by the value of the virtual die. With a 1/6 probability A will receive \$0, and with a 5/6 probability A will receive \$10.

Stage 4 (Results): Finally, A and B are informed of their payoffs. We follow Charness and Dufwenberg (2006) in allowing A to only see her monetary payoff, but not B's roll decision. If A receives a payoff of \$0, she cannot distinguish between B choosing not to roll, or B choosing to roll but getting a bad outcome. This ambiguity makes it more appealing for B to choose not to be trustworthy, particularly when the costs of trustworthiness is high (i.e. $X = \$4$).

In the Renegotiation treatment subjects are described the game as above. However, there is an additional surprise Stage 2.5 after the value of X is realized, but before Player B chooses whether to roll. We make the renegotiation stage a surprise so that subjects' play in Stages 1 and 2 will be the same between the Renegotiation treatment and the baseline. Differences in behavior in later stages will come entirely from the renegotiation itself, rather than differences in the initial promise or trust behavior in Stage 1. [11] The renegotiation opportunity only

[11] While in principle the impact of the outcomes of a surprise versus anticipated renegotiation on the norm for Player A to be trustworthy could be somewhat different, we do not anticipate that such differences would be large in our setting. Recall that our basic intuition is that subjects will react positively to an accepted renegotiation, and negatively to a rejected renegotiation, such that the norm for trustworthiness is strengthened/weakened. If we think about how this might play out differently for an anticipated renegotiation, we would expect that if anything the effect would be even larger, as Player B's may react more negatively to a Player A who refuses a renegotiation that they knew was coming (particularly if that was part of the decision to trust and/or promise. Similarly, the surprise renegotiation may create a bit more uncertainty about the norm (if subjects are not sure how others will react to the surprise). Given this, an anticipated renegotiation may affect subjects propensity to trust or promise, however our main research question is about how the renegotiation outcomes affect the norm to be trustworthy. Hence our design of a surprise renegotiation is cleaner in terms of keeping the trust and promise rates as similar as possible between treatments, and may be somewhat of a conservative underestimate of the impact of renegotiation outcomes on the social norm. We thank Martin Dufwenberg

arises if trustworthiness is costly (i.e. if X is less than 10) - otherwise there is nothing to renegotiate.

Stage 2.5 (Renegotiation): In the Renegotiation treatment when $X = \$4$ or $X = \$7$, i.e. when trustworthiness is costly, B can propose new payoffs to A such that A takes on (in expectation) half the cost of the negative shock.[12] A then has the opportunity to Accept or Reject the request and then B is faced with the Roll/Don't Roll decision. Accepting renegotiation only changes the payoffs when B chooses Roll. If A refuses to renegotiate, the game continues with the payoffs unmodified. Since Player A has no opportunity to punish Player B, and refusing to renegotiate creates the same material payoffs that result when B does not ask for renegotiation, there is no material or strategic reason to forgo requesting to renegotiate. However, we argue that the outcome of the renegotiation may influence the strength of the social norm to follow one's promise.

From a standard neo-classical perspective, play of this game is straightforward. B strictly prefers not to roll if $X = \$4$ or $\$7$, and is indifferent to rolling if $X = \$10$. Additionally, in the standard model the Promise decision is both cheap talk and strategically meaningless, so it will not impact A and B will be indifferent between promising or not. A's can at best hope that B's will roll if $X = \$10$, which only happens 20% of the time, implying an expected value for choosing In of \$1.67. Therefore A's will choose Out (obtaining a payoff of \$2). In the Renegotiation treatment, B is indifferent between asking for renegotiation or not, and will always choose not to roll whether A accepts or rejects. Consequently, A is indifferent between accepting and rejecting, since it will not affect B's behavior. In short, the predictions can be summarized as:

- B strictly prefers not to roll if $X = \$4$ or $\$7$.
- B is indifferent to rolling if $X = \$10$.
- B will be indifferent between promising or not.

for suggesting this discussion point.

[12] In the experiment, the option is presented as the decision to send a (pre-composed) message which is the same for every subject.

- A's will choose Out.
- B is indifferent between asking for renegotiation or not
- A is indifferent between accepting renegotiation or not
- Renegotiation outcomes will have no influence on B's decision to roll.

III. Theoretical Framework

A. Setup

In this section we describe a framework in which decision makers wish to comply with social norms. This is, in some sense, a non-standard conceptualization of utility where the actor's preferences depend on beliefs about collective beliefs regarding appropriate and expected behavior in a particular situation - i.e. preferences depend on the norm.[13] We will first characterize what features any reasonable belief about the norm should have in this setting, and then identify places where individual heterogeneity in beliefs about the norm (e.g. whether there is a strong or weak norm in a particular situation) leads to different actions. We assume that Player A only cares about the monetary payoffs $V(\cdot)$ given the profile of actions taken by each player $\mathbf{a} = (a_A, a_B)$ and the state of nature X:[14]

$$U_A(\mathbf{a}|X) = V_A(\mathbf{a}|X),$$

Player B's utility depends on two components: the monetary payoff V, and the

[13] This idea has been advanced in multiple papers elsewhere. For example, Akerlof and Kranton (2005) note that "...much of utility depends not only on what economists normally think of as *tastes*, but also on *norms* as to how people think that they and others *should* behave...views as to how people should behave depends upon the particular *situation*...".

[14] The alternative would be to have A also have a norm-adherence payoff, like Player B. Having A care about norms would require introducing separate norms governing A's actions (e.g. the decision to trust), which would add complexity and is not necessary to explain A's behavior. The most natural assumption would be that trusting is more normatively appropriate than not trusting, which would just reinforce the prediction that most A's will trust. Additionally, in principal agent settings it is the last-mover (typically the agent) whose behavioral preferences are most important. It is common to assume that the principal has standard preferences, while the agent has behavioral preferences (e.g. Englmaier and Leider, 2012; Dur, Non and Roelfsema, 2010).

value placed on adhering to the norm associated with i's action given the context.[15] The norm-compliance utility can depend on the message sent M (Promise or not) and the state of nature X.

$$U_B(\mathbf{a}|M,X) = V_B(\mathbf{a}|X) + \eta_i N(a_A|M,X),$$

$N(\cdot)$ is a function that assigns to each action a degree of appropriateness or inappropriateness in situation (M,X) undertaken by individual i. In Krupka and Weber (2013) they show that this function can be empirically proxied such that if there is collective recognition that an action constitutes "appropriate," or socially prescribed, behavior, $N(\cdot) > 0$, while if there is joint recognition that an action constitutes "inappropriate", or socially proscribed, behavior, $N(\cdot) < 0$. In their experiments, $N(\cdot) \in [-1, 1]$. These norms vary at the group level and are exogenous at the individual level. Below we will allow that individuals may have differing beliefs about some aspects of N - specifically that everyone agrees which action is more appropriate, but may disagree on the magnitude. The η_i term reflects the degree to which person i cares about complying with the social norms.

In our setting the norm captures the appropriateness of B's final action: the decision to Roll (denoted R) or Don't Roll (denoted D). We assume that the social norm can be different after having sent the message Promise (denoted P) or No Promise (denoted NP), or if the costliness of rolling X is large or small.[16] We can say that there is a social norm for being worthy of A's trust and choosing Roll in a given context if $N(R|M,X) > N(D|M,X)$.

In particular, B's decision will depend on the difference between these norm-compliance utilities $\Delta N = N(R) - N(D)$. Because there are only two actions, it is intuitive that when ΔN is larger, a norm-sensitive individual has a stronger

[15] We start with the assumption that individuals care about behaving in a manner consistent with norms rather than developing a theory of norm compliance based on underlying preferences (Bénabou and Tirole, 2011; Andreoni and Bernheim, 2009).

[16] When clear, for brevity we may suppress the message and/or the state of nature arguments from the norm function.

inclination to choose Roll, hence we can call ΔN the *strength* of the norm.

Definition (Norm Strength): Larger values of $\Delta N(\cdot)$ in the risky trust game represent a "stronger" norm for rolling.

When we add the renegotiation stage, we want to allow that the outcome of the renegotiation may influence behavior through increasing or decreasing the strength of the norm. We will distinguish these cases by adding a superscript to the norm function. We will denote the norm for rolling if there is no renegotiation as $N^{Base}(\cdot)$, the norm when renegotiation is accepted as $N^{Acc}(\cdot)$, and the norm when renegotiation is rejected as $N^{Rej}(\cdot)$.

B. Assumptions for the Norm in the Baseline Game

We are now in a position to make the following assumptions about the social norms function. Our first assumption is motivated by empirical evidence that B's choose to roll even when there is no promise or communication in these games. Charness and Dufwenberg (2006) observe in their (5,5) treatment without B messages that 44% of B's chose to roll. Based on these observations, we make the assumption that there will be a norm to roll for any realization of the state of nature X.

Assumption 1 (Norm of trustworthiness): For each value of X, Rolling is seen as more appropriate.

$$\forall X \ \Delta N(X) > 0$$

Assumption 2 (Promises change the norm): For each value of X, Promises increase the normative prescription to roll.

(a) $\forall X \ N(R|P,X) > N(R|NP,X)$ and $N(R|P,X) > 0$

(b) $\forall X \ N(D|P,X) < N(D|NP,X)$ and $N(D|P,X) < 0$

(c) $\forall X \ \Delta N(P,X) > \Delta N(NP,X)$

Our second assumption is motivated by recent findings showing that informal promises have a large effect on behavior(Charness and Grosskopf, 2004; Fehr, Hart and Zehnder, 2011; Feltovich and Swierzbinski, 2011; Miettinen, 2013; Krupka, Leider and Jiang, 2016), even when fulfilling the promise entails a personal cost (Ellingsen and Johannesson, 2004; Charness and Dufwenberg, 2006; Vanberg, 2008; Kessler and Leider, 2012; Dufwenberg, Gächter and Hennig-Schmidt, 2011). This literature also shows that making a promise elicits a promise-keeping norm which is strongly proscriptive (Krupka, Leider and Jiang, 2016). This suggests that making a promise strengthens the norm to follow that promise and roll (this corresponds to (a) in assumption 2). We also expect that not rolling after making a promise is (more) inappropriate (this corresponds to (b) in assumption 2). Moreover, while the size of this effect may vary depending on the cost of rolling, we expect that promises will have a norm-strengthening effect for any state of nature (this corresponds to (c) in assumption 2).

Taken together, assumptions one and two characterize features of the norm that, given prior empirical findings on behavior, we expect to find empirical support for when we elicit beliefs about the norms in our experiments. Specifically, they constitute our hypotheses regarding the norm function.

Hypothesis 1 (Norm of trustworthiness): Subjects will rate the action Roll as more appropriate than Don't Roll, for all states of X.

Hypothesis 2 (Promises change the norm): The difference in subjects' appropriateness ratings between Roll and Don't Roll will be larger when B made a promise than when B did not make a promise, for all states of X.

The assumptions about the norm also allow us to make specific predictions about behavior. We can predict that making a promise will increase the likelihood of rolling if there is a (non-empty) interval of values for η such that B will choose to roll if he has promised, and will choose not to roll if he has not promised. This means that there will be some subjects who will behave differently depending on

whether there is a promise.

Using the assumptions about the norm, we can see that such an interval exists. If B chooses to roll, he receives $U(R) = X + \eta N(R)$, while if he chooses not to roll, then he receives $U(D) = 10 + \eta N(D)$. Therefore, he will be willing to roll if the increase in his norm utility from rolling outweighs the decrease in his monetary payoff, i.e. if $\eta \Delta N \geq (10 - X)$. This is always true if $X = \$10$. For $X = \$7$, $\eta \in [3/\Delta N(P, 7), 3/\Delta N(NP, 7)]$.[17] Similarly, for X = \$4 the promise would change behavior if $\eta \in [6/\Delta N(P, 4), 6/\Delta N(NP, 4)]$. Given our assumptions about the norm, in both the case where $X = \$7$ or \$4, the intervals are non-empty and so we should expect to see that making a promise increases rolling.[18]

B's will promise if the increased normative utility from rolling outweighs any decreased monetary utility from a change in behavior.[19] Using our assumptions about the norm, we can see that such an interval exists. To see this, suppose we fix X. Then a Player B can choose to Promise and Roll, receiving $U = X + \eta N(R|P)$, or Not Promise and Don't Roll, receiving $U = 10 + \eta N(D|NP)$. He will then make the promise if $\eta \times (N(R|P) - N(D|NP)) > 10 - X$. Since $\Delta N(P) > N(R|P) - N(D|NP) > \Delta N(NP)$ it is possible to find such a value of η for any X, given our maintained assumptions.

However, B's choose to promise before they know what X is. Therefore, they will only be willing to promise if his expected utility across the states of promising is greater than the expected utility of not promising. There are many possible values of η and $N(\cdot)$ that are consistent with B's promising, so we discuss here one example to demonstrate that a social norms model can justify promise making in our setting. Suppose Player B anticipates that, given the norms, he will roll for any realization of X if he made the promise, and that if he did not make the

[17] To see why this is true, note that $\eta \Delta N(P, 7) \geq 3$ and $\eta \Delta N(NP, 7) < 3$.

[18] The intuitive interpretation is that promises will make a difference in B's behavior if η takes on an intermediate value such that the strong norm under a promise is sufficient incentive to roll, but the weak norm without a promise is not sufficient.

[19] If a Player B is willing to roll even without a promise, then making a promise is straightforward. Similarly, if Player B does not care about the norm (i.e. if $\eta = 0$) then the promise is just cheap talk and he is happy to make it. We are interested in cases where a player is willing to make the promise when doing so will change his behavior for at least one value of X.

promise he would roll if $X = \$7$ or $\$10$ but not if $X = \$4$. Then to be willing to choose Promise he needs to have $U(Promise) > U(NoPromise)$, i.e.:

$$\begin{aligned} 0.2 \times (\eta[N(R|P,10) - N(R|NP,10)]) + \\ 0.4 \times (\eta[N(R|P,7) - N(R|NP,7)]) + \\ 0.4 \times (\eta[N(R|P,4) - N(D|NP,4)] - 6) \geq 0 \end{aligned}$$

The first two terms are clearly positive, so a sufficient condition is for the last term to be positive.[20] It must also be the case that he wold roll if $X = \$4$ if he promised, and would not roll if he did not promise. By assumption 2, we obtain that $\eta \times \Delta N(P,4) \geq 6$, and $6/\Delta N(NP,4) \geq 6/(N(R|P,4) - N(D|NP,4))$. Therefore, there is an interval where if $\eta \in [6/(N(R|P,4) - N(D|N,4)), 6/\Delta N(\cdot|N,4)$, then the player would be willing to make a promise.[21]

Finally, A chooses In if she believes sufficiently many B's will Roll. If A chooses Out she will get \$2, while if she choose In she will get (on average) \$8.33 if Player B rolls, and \$0 if Player B doesn't roll. Therefore, A will be willing to choose In as long as the probability that B rolls is at least 24%. Therefore, A would be willing to choose In as long as enough Player B's will make the promise and the concern for norms is sufficiently strong.[22]

Thus, using our assumptions about the norm, we can make the following specific predictions about behavior in the Baseline game:

Hypothesis 3 (Player A behavior: In): Some A's will choose In.

Hypothesis 4 (Player B behavior: Promise making): Some B's will Promise.

Hypothesis 5 (Player B behavior: Promise & Roll): Making promises will

[20] That is, we want $\eta[N(R|P,4) - N(D|NP,4)] \geq 6$, or $\eta \geq 6/(N(R|P,4) - N(D|NP,4))$

[21] The intuitive interpretation is that B's will promise if the increased normative utility from rolling outweighs any decreased monetary utility from a change in behavior.

[22] Different combinations of B players (and their associated values of η) could support this belief - both those that will roll for every value of X, and/or those that roll only for some X realizations.

increase the fraction of B's who Roll.

To test these predictions about our baseline, we run an experiment.

IV. Experiment: Design

Our experimental design consists of two treatments (Baseline and Renegotiation) that will collect data to test our hypotheses regarding both behavior and norms (in the Appendix, we provide a diagram of the experiment design). Each subject participates first in either the Baseline or Renegotiation risky trust game and then in the norm elicitation task; the former collects data on choices and the latter on beliefs about the norm. The risky trust game is different depending on which treatment a subject is in but the norm elicitation task does not vary between treatments and elicits beliefs about the norms for the Baseline and Renegotiation games. Both the risky trust game and the norm elicitation task are incentivized to yield a combined final payoff.

Table 1 depicts a summary of choices and associated payoffs for the Baseline risky trust game. These choices and payoffs are common knowledge, as are the probabilities: p(X=\$4) is 0.4; p(X=\$7) is 0.4; p(X=\$10) is 0.2.

Table 1—Payoff table for Risky Trust Game in Baseline Treatment

Baseline	**A Earns**	**B Earns**
A chooses Out	\$2	\$2
A chooses In, B chooses Don't Roll	\$0	\$10
A chooses In, B chooses Roll, die=1	\$0	\$X
A chooses In, B chooses Roll, die=2,3,4,5,6	\$10	\$X

In the Renegotiation treatment, after X is determined (to be either \$7 or \$4) but before the roll decision, B is informed that he has a chance to propose new payoffs before making the Roll decision. B has the option of sending Player A a pre-written message that proposes to change the payoffs to those depicted in

Table 2.[23] If A accepts, then both players are informed that it was accepted and then B makes a Roll decision. If B does not send the message or if A rejects the proposed changes, then the Baseline payoffs are preserved, A and B are informed that the proposal was rejected, that payoffs remain unchanged (i.e. they are those depicted in Table 1), and then B makes a Roll decision.

TABLE 2—PAYOFF TABLE AFTER ACCEPTED RENEGOTIATION

Renegotiated payoffs if X=$4	**A Earns**	**B Earns**
A chooses Out	$2	$2
A chooses In, B chooses Don't Roll	$0	$10
A chooses In, B chooses Roll, die=1,2,3	$0	$7.00
A chooses In, B chooses Roll, die=4,5,6	$10	$7.00
Renegotiated payoffs if X=$7		
A chooses Out	$2	$2
A chooses In, B chooses Don't Roll	$0	$10
A chooses In, B chooses Roll, die=1,2	$0	$8.50
A chooses In, B chooses Roll, die=3,4,5,6	$10	$8.50

Regardless of whether subjects are in the Baseline or Renegotiation treatment, the next task subjects complete is the norm elicitation task.[24] Subjects first read a vignette that describes the Baseline. After reading the vignette, subjects are asked to evaluate how appropriate rolling and not rolling is for each unique combination of X values and whether B has made a promise or not. The unique feature of this protocol is that they are asked to coordinate with another subject on the rating that "most other people in the room believe constitutes socially appropriate or inappropriate behavior".

[23]The message text for X=$4, for example, is: "Hi. Now that I know that X=4, I want to ask if you would agree to change the dice game if I choose to roll. I would like to propose that we change the payoffs such that you get $0 and I get $7 if the die roll comes up 1, 2 or 3. If the die roll comes up 4, 5, or 6 you get $10 and I get $7."

[24]Krupka and Weber (2013) developed the norm elicitation protocol as an empirical proxy for the norm. It uses coordination games to identify which actions are jointly recognized to be socially appropriate or inappropriate. Krupka and Weber (2013) show that collectively-recognized social norms create focal points in the matching game. If there is a social norm that some actions are more or less socially appropriate, respondents attempting to match others' appropriateness ratings are likely to rely on this shared perception to help them do so. Thus, the incentive in the coordination game elicits collective perceptions of appropriateness which we will call our empirical proxy of the social norm.

For example, after reading the baseline vignette,[25] they are prompted with a list of possible actions available to B and asked to play the coordination game over appropriateness ratings for each action. The appendix provides a figure which depicts two of the actions that subjects rate and provides an example of how X is varied while all other aspects are held constant.

Subjects receive an incentive to match their appropriateness rating with that of another, randomly selected, subject (who is reading the same vignette and in the same session). They can choose between characterizing the Roll decision as anything from "very socially appropriate" to "very socially inappropriate". We use a proper scoring rule such that each category by which they differ from their matched counter-part, means the subject looses \$4. Thus, if they are off by one appropriateness category in either direction, then they are paid \$11 and so on.[26]

The norm elicitation task yields three measures that we will use. We obtain a subject's beliefs about the norm from his/her individual responses in the coordination game. We have an empirical proxy of the norm from the mean ratings in the coordination game. We have an empirical proxy for norm strength which we obtain from changes in mean ratings as we vary different features of the context (eg. the value of X).

A total 282 subjects participated in the experiment with 94 in the Baseline and 188 in the Renegotiation treatment. The average total payment from the risky trust game and the norm elicitation task in the Baseline condition was \$16.74 and in the Renegotiation treatment was \$16.97.

[25] For subjects in the Baseline, the baseline vignette will describe the situation they just played, but for those in the Renegotiation treatment, the baseline vignette will present a novel situation. The reverse will be true when subjects read the Renegotiation vignette.

[26] To incent subjects to think about what others think is appropriate, we used a proper scoring rule (Lambert and Shoham, 2009) to elicit subjects' median belief about the distribution of others' ratings by matching a subject with another subject and then paying them according to the following payoff function: n_i=\$15-\$4$|x_i - x_{-i}|$, for each subject i. Where n_i is the payoff of subject i, and x_i and x_{-i} are the appropriateness ratings for subject i and the matched other subject, respectively. In short, this scoring rule pays them \$15 if their appropriateness rating matches that of another randomly selected subject's appropriateness rating and we subtract \$4 for each additional category above or below that they are off from their match's rating.

V. Results: Baseline

Our hypotheses about behavior in the baseline of the risky trust game are motivated by two assumptions that characterize features of the norm. We use data from the baseline vignette in the norms elicitation task to test the hypotheses associated with those assumptions.[27]

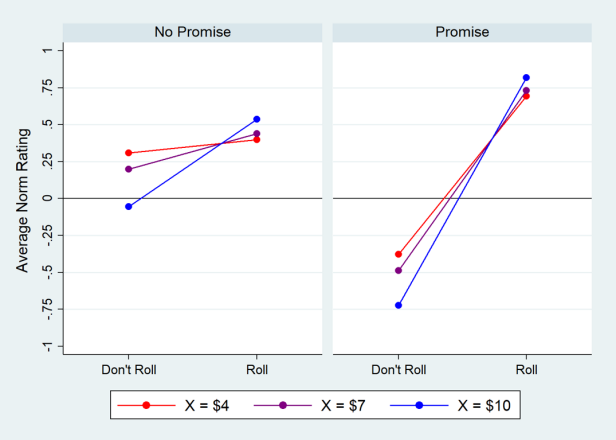


FIGURE 2. AVERAGE NORM RATINGS FOR THE BASELINE.

Figure 2 depicts the average norm ratings in the baseline vignette. On the left panel, where B has not made a promise, the average norm ratings for Roll are always higher than for Don't Roll. On the right panel, where B has made a promise, Roll is considered very appropriate and Don't Roll is inappropriate. For both the left and right panels, the differences in average norm ratings between Don't Roll and Roll differ significantly (signed rank test $p < 0.01$ for both) and are consistent with Hypothesis 1 (a norm of trustworthiness). In addition, there is support for Hypothesis 2 (promises change the rolling norm). The change in the

[27] We convert subjects' norm ratings into equally spaced numerical scores. A rating of "very socially inappropriate" is assigned a score of -1, "socially inappropriate" a score of -0.6, "somewhat socially inappropriate" a score of -0.2, "somewhat socially appropriate" a score of 0.2, "socially appropriate" a score of 0.4 and "very socially appropriate" a score of 1. This transformation follows Krupka and Weber (2013) and is discussed there. It is also used in numerous other studies: Kimbrough and Vostroknutov (2016), Gächter, Nosenzo and Sefton (2013), Veselỳ (2015), Erkut, Nosenzo and Sefton (2015), d'Adda, Drouvelis and Nosenzo (2016), Gangadharan et al. (2016), Banerjee (2016), and Gächter, Gerhards and Nosenzo (2017).

norms ratings when there is a promise is larger than the change in norms ratings when there is no promise. Put differently, $\Delta N(\cdot|P) > \Delta N(\cdot|N)$ is significant at $p < 0.01$ for all ΔN comparisons using the signed rank test.

Given that we find support for our hypotheses about the social norms, we now turn to the behavior that such norms would predict. Though our focus in this section is on the Baseline, we pool the Renegotiation treatment and Baseline to report A's In/Out choice and B's Promise choice because the treatment does not differ from the Baseline at this stage.

We find that 76% of A's decide to go In.[28] We also find that 82% of B's Promise.[29] These results are consistent with Hypothesis 3 (Player A behavior: In) and Hypothesis 4 (Player B behavior: Promise making).

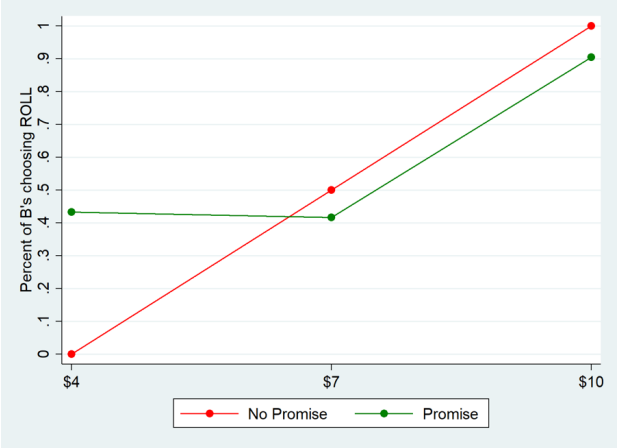


FIGURE 3. FRACTION OF B'S CHOOSING ROLL IN THE BASELINE.

Figure 3 shows the Roll decision as a function of whether a promise was made and the value of X as either \$4, \$7 or \$10. With or without a promise, the probability of rolling decreases with lower values of X. Rolling when $X = \$4$ is 32%, when $X = \$7$ is 43% and when $X = \$10$ is 92% (non-parametric trend test;

[28] 72% in the Baseline and 78% in the Renegotiation treatment; proportions test diff. n.s. between treatments.

[29] 83% in the Baseline; 82% in the Renegotiation treatment; proportions test diff. n.s.

$p < 0.01$). However, choosing Promise increase rolling when $X = \$4$ (43% vs. 0%; proportions test; $p < 0.01$), but they have no effect when $X = \$7$ or $X = \$10$ (42% vs. 50%, 100% vs 90%; proportions test; diff. n.s.). A regression (reported in the appendix) also finds that Promise significantly differs depending on the value of X (the marginal effect of the promise on rolling for $X = \$4$, \$7 and \$10: 1.55, $p < 0.01$; -0.063, diff. n.s.; -1.22, $p < 0.01$). These results are consistent with Hypothesis 5 (Player B behavior: Roll with a promise).

Our results for the Baseline treatment are in line with the behavior we expected given our assumptions about the norms. Using the Renegotiation treatment, we identify additional specific predictions about norms and behavior that would add further support to using the norms framework.

VI. Renegotiation Treatment: Predictions and Aggregate Results

We begin by advancing several assumptions regarding the impact of an accepted or rejected renegotiation request on the norm and behavior. We then test those assumptions and the predicted behavior patterns. In doing so, we first consider aggregate patterns, and then identify several individual patterns drawing connections between behaviors and beliefs about the norm.

Recall that after an accepted renegotiation A agrees to take on roughly half of the negative shock. A large body of evidence regarding norms of reciprocity informs our intuitions on how rejection or acceptance of a renegotiation may impact the strength of the norm in our setting (Fehr and Gächter, 2000, 1998). This literature suggests that if A accepts a renegotiation, it will increase the imperative to roll and make it more inappropriate not to roll. Thus, acceptance will increase norm strength, $\Delta N^{Acc}(\cdot)$. Similarly, if A rejects a renegotiation request, it decreases the difference between how appropriate Roll and Don't Roll are. We therefore make the following assumption about how norm strength is affected by the renegotiation:

Assumption 3 (Norm strength: Renegotiation): If A accepts renegotiation, it strengthens the norm of rolling relative to the baseline; while rejecting the renegotiation weakens the norm relative to the baseline, for any state of nature X.

Assumption 3 produces a straightforward hypothesis about the norms ratings in the Renegotiation treatment:

Hypothesis 6 (Norm strength: Renegotiation): Norm strength, ΔN, will be larger than the baseline after an accepted renegotiation and smaller than the baseline after a rejection, for any state of nature X.

$$\Delta N^{Acc} > \Delta N^{Base} > \Delta N^{Rej} \ \forall X$$

Assumption 3 implies that the decision to renegotiate is not straightforward for a norm-sensitive individual. Though there is no strategic reason for B to forgo requesting to renegotiate, asking for renegotiation has a down side.[30] If B asks for renegotiation, there is some probability that A accepts and B gets the benefit of the renegotiated terms and a boost to the norm utility. But it is also possible that B gets rejected. The rejection would reduce the norm utility: Roll becomes less appropriate and Don't Roll more appropriate relative to the Baseline or to the accepted request. For some B's, the reduction in norm utility may be large enough to keep them from asking in the first place.[31] Thus, we predict:

Hypothesis 7 (Some Player B's won't ask for renegotiation): *There will be B subjects who do not ask for renegotiation.*

Table 3 reports the average norm strength (ΔN) when a renegotiation was accepted (most right column) or rejected (most left column) and in the baseline (middle column). Hypothesis 6 (Norm strength: Renegotiation) finds broad

[30] Recall that in our game, A has no opportunity to punish B, and refusing to renegotiate creates the same material payoffs as obtained if Player B does not ask for renegotiation.

[31] If A's anticipate that B will roll regardless of how renegotiation works out, then A is free to reject and it justifies B's fear.

support. When X=\$4 the norm strength is greater after an accepted request for renegotiation than the baseline ($0.46 > 0.09$ when no promise was made; $1.37 > 1.07$ when a promise was made). It is weaker than the baseline after a rejection ($0.09 > -0.10$ No Promise; $1.07 > 0.46$ Promise). Similar patterns hold when X=\$7 (signed rank test $p < 0.01$ for all pairwise comparisons for both $X = 4$ and $X = 7$).

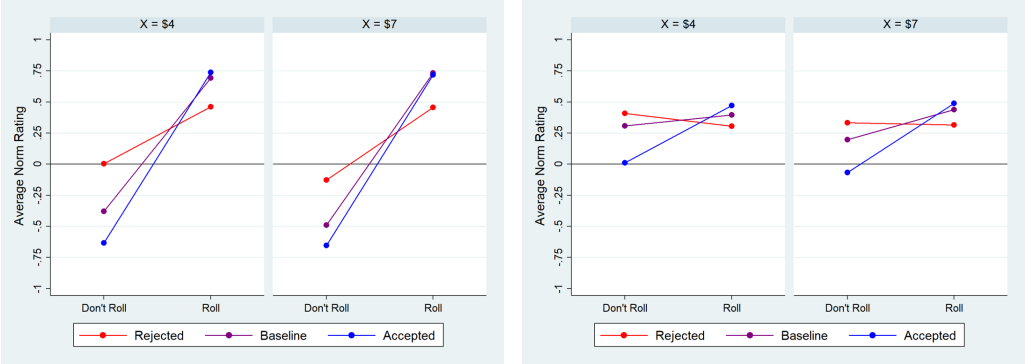
Table 3—Change in Norm Strength.

	X=\$4			X=\$7		
	ΔN^{Acc}	ΔN^{Base}	ΔN^{Rej}	ΔN^{Acc}	ΔN^{Base}	ΔN^{Rej}
No Promise	0.46	0.09	-0.10	0.55	0.24	-0.02
Promise	1.37	1.07	0.46	1.37	1.22	0.58

Table 4 is an alternative way to visualize table 3 and depicts the mean norm ratings for the Roll and Don't Roll action where X is either \$4 or \$7. It is possible to see that the difference between average norm ratings for rolling and not rolling when a renegotiation is accepted (blue line) is larger than the baseline and larger than when it has been rejected (red line). This is true for both situations where a promise was (top panel) and was not made (bottom panel).

As expected, the elicited social norms create a downside (in utility terms) for a failed renegotiation. For example, when X=\$4, a Player B who will choose Roll even if rejected risks a substantial decrease in the norm utility (1.07 vs. 0.46) if renegotiation is rejected. In support of Hypothesis 7 (Some Player B's won't ask for renegotiation), we see many subjects choose not to renegotiate. When X=\$4, 45% of B subjects do not ask and when X=\$7, 68% of B subjects do not ask. Though A's should accept (on average, A's who accept get about 44% more money), 45% reject when X=\$4 and 50% reject when X=\$7. Thus, asking is risky for B's and, if B's anticipate some chance of rejection, it explains why some B's don't ask.

Having shown that features of observed data in the risky trust game are con-



(a) Promise (b) No Promise

FIGURE 4. AVERAGE NORM RATINGS FOR RENEGOTIATION.

sistent with the norm we identified, we offer a third analysis in which we exploit individual beliefs about the norm to provide additional support for our framework.

VII. Renegotiation Treatment: Individual Heterogeneity

Up until this point, our analysis has used the average ratings in our coordination game to create a proxy for the norms function.[32] However, subjects also exhibit variation in their beliefs about the norm.

There are a variety of individual beliefs about the norm that are consistent with the norms function we characterized in our main analysis. Subjects who agree about the relative norm strength across renegotiation outcomes (i.e. the inequality laid out by Assumption 3), may differ in the magnitude of the difference.[33] We show how these different beliefs about the norm lead to different patterns of behavior for B's. We want to explain B's decision both to roll (or not) after different renegotiation outcomes, as well as the decision to ask for renegotiation based on their individual norm belief profiles.

[32]Individuals can differ in their sensitivity, captured by our parameter η, or in their beliefs about the social norm, N. In our experiment we are not able to directly measure an individual's η, but do observe their elicited belief about the norm.

[33]For example, if we restrict ΔN^{Rej} to be less than ΔN^{Base}, it could be that there is heterogeneity in whether an individual believes that ΔN^{Rej} is large or small in magnitude.

The decision to choose Roll is straightforward given a specific norm belief profile, however the decision to ask for renegotiation also depends on B's belief about A's willingness to accept. We take a narrow approach to motivate our data analysis that leverages the fact that B's social norm beliefs reflect B's understanding about how *most people* will view this game.[34]

First, recall that we have assumed that A's only care about their material payoff. Materially-focused A's will only accept a renegotiation request if it will change B's action (i.e. that B will roll if accepted and will not roll if rejected). To pin down B's beliefs about whether A will accept, we assume that B's beliefs about A's action will be consistent with B's beliefs about the norm. For example, if B's project their norm beliefs onto everyone, then a specific B subject will be optimistic about A's accepting a request if acceptance would change that B's Roll behavior. By similar logic, B will be pessimistic about A's accepting if B would take the same action whether A accepts or rejects. Of course, B's that ask may be surprised by A's decision to accept or reject. This will be empirically useful for us, as some of these "surprise" outcomes will allow us to separate belief types.

Our approach is therefore to take each belief profile type, describe the implied (unobservable) belief about A's likely acceptance, and then identify the (observable) decision to ask and a set of (potentially observable) Roll decisions. This generates a set of testable predictions about how various sets of behaviors should be associated with different norm belief profiles.

Table 4 depicts four norm belief profiles and their associated behavior predictions. "SSS" denotes a norm belief profile of a subjects who believes that there are strong norms when a renegotiation request has been accepted, strong norms in the baseline, and strong norms even when a renegotiation request has been rejected. "SSw" denotes a norm belief profile of a subjects who believes that there are strong norms when a renegotiation request has been accepted, strong

[34] A more general approach would involve a full equilibrium model where A's and B's have distributional beliefs over the other players' type that reflect the population shares. While interesting, such a model is beyond the scope of this paper.

norms in the baseline, and *weak* norms when a renegotiation request has been rejected. While "www" denotes a norm belief profile of a subjects who believes that there are weak norms when a renegotiation request has been accepted, weak norms in the baseline, and weak norms when a renegotiation request has been rejected. "Mww" reflects a belief in a moderately strong norm after acceptance which we will describe below.

Table 4—Summary of Norm Belief Cases.

			Roll if...		
Profile	Beliefs re: A	Will ask?	D/n ask	Acc	Rej
SSS	Pessimistic	No	Yes	-	-
SSw	Optimistic	Yes	-	Yes	No
www	Pessimistic	Yes	-	No	No
Mww	Optimistic	No	No	-	-

Note: Column 1 lists four norm profiles that predict specific patterns of behavior. Column 2 lists the associated beliefs about Player A's likelihood of accepting (given our assumption about Player B's projecting their norms onto others). The remaining columns list the implied Ask and Roll behavior. "SSS" denotes a subject with a strong norm if renegotiation is accepted, a strong norm in the baseline, and a strong norm if rejected. "SSw" subjects' profiles differ in that they have a weak norm if rejected. "www" subjects believe that the norm is weak in all cases, while "Mww" subjects believe that the norm after acceptance is moderately strong.

Our first case considers subjects who think that there is a strong norm for any renegotiation outcome. They have a norm profile denoted SSS in table 4. While we expect that in general a rejected renegotiation will in some way weaken the norm, these subjects believe that it does not weaken it very much. With this profile, B would not ask for renegotiation if they anticipate that asking would lead to being rejected and rolling anyways, since rolling after rejection leads to lower norm utility than rolling in the base case. This fear of rejection is justified by B's belief that the social consensus is for a strong norm even after rejection: if B would still choose Roll if rejected then A materially benefits from rejecting the request. Hence a B who will roll even if rejected will, under our belief assumption, expect that A's will reject a renegotiation request. In the Appendix we show that this

scenario is most likely to occur if η is large and/or all ΔN (but most importantly ΔN^{Rej}) is large. For these strong-norm subjects, they prefer simply to roll rather than risk the lower norm utility for rolling after rejection. We predict:

Hypothesis 8 (All Strong Norms implies Don't Ask, Roll): B subjects who do not ask for renegotiation and then roll anyway will have individual beliefs about the norm that satisfy: ΔN^{Rej} is large compared with ΔN^{Rej} of other subjects and ΔN^{Base} and ΔN^{Acc} to be large.

In order to test Hypothesis 8 with our norm and behavioral data, we compare both ΔN^{Base} and ΔN^{Rej} between subjects with different patterns of behavior. We expect to see that subjects who do not ask for renegotiation and then roll will have a very high average for both ΔN^{Base} and ΔN^{Rej}.

Our second profile describes another intuitive pattern: that rejection results in a large reduction in the norm strength such that the roll norm after rejection is *weak*. We can think of these subjects as having a norm profile denoted SSw in table 4. In contrast to the first case, these subjects have less downside from asking for renegotiation, as the weakened norm can allow them to not roll after a rejection. Additionally, such B's believe that A's will anticipate this difference in behavior from rejection, hence the B players will expect that renegotiation will be accepted, further increasing the upside to B asking for renegotiation. Hence, this belief profile is associated with B's asking for renegotiation and then rolling if the request was accepted.

Therefore, we expect to see that B's who ask and then choose Roll are those who believe that there is a strong norm to roll in the baseline, a strong (or stronger) norm to roll after acceptance, and who believe that the norm is significantly weakened when A has rejected the request. In the Appendix we identify a non-empty interval of η values that are sufficient for this condition to obtain. Here we simply note the prediction:

Hypothesis 9 (Strong Norms Weakened by Rejection implies Ask, Roll):

B subjects who asked for renegotiation (were accepted) and then rolled, will have individual beliefs about the norm that satisfy $\Delta N^{Acc} \geq \Delta N^{Base} > \Delta N^{Rej}$.

Based on Hypothesis 9 we expect to see that subjects who ask for renegotiation, and who roll when accepted, will have a very low average ΔN^{Rej} and a very large average ΔN^{Base} and/or ΔN^{Acc} , reflecting the substantial weakening of the norm from a rejected renegotiation request and the overall strength of the norm in other cases.

A third profile would be that the subjects believe the norm is relatively weak in all cases. We can think of them as having a norm profile denoted "www" in table 4. These subjects will act most like the standard prediction - they see little downside to asking for renegotiation, even if they expect rejection. However, focusing on the B's in our data who do not roll after rejection is not very informative, as a B player with beliefs from the second case above would also not roll if surprised by a rejection. The most distinctive behavioral pattern for subjects with these beliefs is that they will not roll *even if accepted*. No other pattern of norm beliefs is associated with subjects who both ask for renegotiation and do not roll when accepted. In the Appendix we show that this can occur when η is small and/or ΔN is small (in particular we need ΔN^{Acc} to be small). The theoretical prediction is therefore:

Hypothesis 10 (All Weak Norms implies Ask, Don't Roll after acceptance): B subjects who ask, and receive an acceptance, and then don't roll, will have individual beliefs about the norm that satisfy: all ΔN's are small, especially ΔN^{Acc}.

Finally, a B with a moderately strong norm after acceptance may choose not to ask for renegotiation because they anticipate being accepted (which would potentially cange their behavior), but prefer the higher monetary payoff from not rolling to the increased normative payoff of being accepted and rolling. Such a

pattern would be consistent with the norms model if B believes there is a weak norm in the baseline (small ΔN^{Base}), a reasonably strong norm for rolling if a request were accepted (i.e. ΔN^{Acc} is large), but that the increased utility from switching behavior after a potential acceptance isn't that attractive (i.e. $N^{Acc}(R|X) - N^{Base}(D|X)$ is small). We identify these subjects as having a norm profile denoted "Mww" in table 4. We would therefore predict:

Hypothesis 11 (Weak Baseline Norm implies Don't Ask, Don't Roll): Subjects that do not ask for renegotiation and do not roll have beliefs about the norm such that $N^{Acc}(R|X) - N^{Base}(D|X)$ is small, ΔN^{Acc} is large, and ΔN^{Base} small.

We now explore the renegotiation data to look for evidence consistent with these hypotheses. Table 5 reports the average norm strength of B's separated out by what actions B took. Column 1 reports the average norm strength for those B's who did not ask for a renegotiation and rolled. Column 2 reports the norm strength for those B's who did not ask for renegotiation and did not roll. Column 3 reports the norm strength for all B's who asked for renegotiation (regardless of whether they rolled or did not roll). We break column 3 out further in Table 6.

TABLE 5—AVERAGE NORM STRENGTH (ΔN) BY ASK/ROLL BEHAVIOR.

	ΔN for B's by Action		
	Do Not Ask Roll	Do Not Ask Do Not Roll	Ask (pool roll choice)
ΔN^{Acc}	1.47	1.20	1.15
ΔN^{Base}	1.53	0.78	0.96
ΔN^{Rej}	1.27	0.57	0.46

We begin by focusing on subjects who did not request renegotiation (i.e. those described by Hypotheses 8 and 11). In Column 1 we see that subjects who did not ask and rolled have a large ΔN in all three renegotiation scenarios, in line with

Hypothesis 8 (All Strong Norms implies Don't Ask, Roll). The difference is large and significant for ΔN^{Rej}, the key determinant for this pattern, ($\Delta N^{Rej} = 1.27$ vs 0.57 and 0.46; $p = 0.09$ compared to column 2, $p = 0.01$ compared to column 3, $p = 0.02$ compared to columns 2 and 3 pooled). ΔN^{Base} is also substantially and significantly larger for these subjects ($\Delta N^{Base} = 1.53$ vs 0.78 and 0.96; $p = 0.02$ compared to column 2, $p = 0.05$ compared to column 3, $p = 0.03$ compared to columns 2 and 3 pooled). ΔN^{Acc} is directionally larger, but not significantly so ($\Delta N^{Acc} = 1.47$ vs. 1.20 and 1.15).

As an additional measure of overall norm strength across scenarios we can take the sum of the ΔN values (i.e. we sum column 1 along the three cases of accept, reject and baseline). Comparing this sum, we find that the subjects in Column 1 have a much stronger total norm strength (column 1 compared to column 2 is 4.27 vs 2.55, $p = 0.05$; column 1 vs 3 is 4.27 vs 2.58, $p = 0.04$; column 1 compared to columns 2 and 3 pooled, $p = 0.03$). Hence we find evidence supportive of Hypothesis 8: subjects with "SSS" belief profile are likely to *not* ask for renegotiation and roll anyway. This is because these subjects worry about being rejected and receiving less norm utility when they roll anyway.

Column 2 data support Hypothesis 11 (Weak baseline norm implied don't ask, don't roll). The intuition for why these subjects do not ask is that an acceptance would change their behavior, and they prefer the monetary payoff of not rolling to the norm utility. Those who did not ask and did not roll do believe the norm is strong after acceptance: the average $\Delta N^{Acc} = 1.20$ for these subjects, which is larger than $\Delta N^{Base} = 0.78$ (indicating that an acceptance might change their roll decision). Further, 1.20 is comparable to the ΔN^{Acc} of 1.47 and 1.15 from the other columns (Kruskal-Wallis and pairwise rank-sum tests, diff. n.s.).

Consistent with not rolling after not asking, the baseline norm is relatively weak: ΔN^{Base} in column 2 (0.78) is directionally smaller than ΔN^{Base} for column 1 and 3, with a significant difference compared to column 1 (rank sum test, $p < 0.01$). Finally, $N^{Acc}(R|X) - N^{Base}(D|X)$ is 0.93, which is comparable to the average

for subjects that renegotiated (0.99), and directionally smaller than for subjects who Do Not Ask and Roll (1.33). We also find that ΔN^{Rej} is significantly smaller than the corresponding norms for subjects in column 1 (0.57 vs 1.27), which is a natural consequence of ΔN^{Base} being small. These results support Hypothesis 11.

To explore Hypotheses 9 and 10, Table 6 breaks down the ΔN for those B's who requested renegotiation, separated by the outcome of the renegotiation (accept or reject) and B's subsequent Roll decision. Columns 1 and 2 report the average ΔN values for subjects whose renegotiation request was accepted; column 1 reports average ΔN for subjects who ultimately rolled, while column 2 reports those values for subjects who did not roll. Similarly, columns 3 and 4 report the average ΔN values for subjects whose renegotiation request was rejected; column 3 for subjects who rolled, and column 4 for subjects who did not roll.

TABLE 6—AVERAGE NORM STRENGTH (ΔN) AFTER RENEGOTIATION.

	ΔN for B's who asked for renegotiation			
	A accepts		A rejects	
	B rolls	B doesn't roll	B rolls	B doesn't roll
ΔN^{Acc}	1.28	0.80	1.20	1.16
ΔN^{Base}	0.80	0.48	1.27	1.16
ΔN^{Rej}	−0.08	−0.08	1.07	0.87

The subjects in Column 1 are described by Hypothesis 9 (Strong Norms Weakened by Rejection implies Ask, Roll), and as expected they have one of the weakest norms after rejection, and the largest reduction from the baseline norm. For these subjects ΔN^{Base} is 0.80 and ΔN^{Rej} is -0.08. ΔN^{Rej} is significantly smaller than ΔN^{Base} in column 1; it is also smaller than ΔN^{Rej} in columns 3 and 4 (Kruskal-Wallis joint test for differences between all columns is $p < 0.01$, pairwise tests significant between 1 versus 3 or 4, $p < 0.01$ for both).

Similarly, if we compare the norm difference ($\Delta N^{Base} - \Delta N^{Rej}$) to all other

subjects who renegotiated (i.e. pooling columns 2 to 4), we find that column 1 has the largest difference (0.88 versus 0.53, $p < 0.01$).[35] Finally, we also see that ΔN^{Acc} is 1.28. This value is larger than those in columns 3 and 4 and significantly larger than column 2 ($p = 0.05$). These results provide support for Hypothesis 8 that asking and then rolling when accepted should be associated with having strong norms for acceptance and baseline, but weak norms after rejection.

Column 2 gives us the norms for subjects associated with Hypotheses 10 (All Weak Norms implies Ask, Don't Roll after acceptance). As expected we find that ΔN^{Acc} is directionally smaller (0.80) when compared to the beliefs of other B subjects (across columns 1-4 for ΔN^{Acc}), and is significantly smaller than column 1 ($\Delta N^{Acc} = 1.28$ significantly differs from column 2; $p = 0.05$), and marginally significantly smaller than the other three columns (i.e. 1, 3 and 4) when pooled ($p = 0.09$). We also find that $\Delta N^{Base} = 0.48$ is smallest for column 2 subjects, and significantly differs compared to column 3 (1.27), to column 4 (1.16) and to the columns 1 and 3-4 pooled ($p < 0.05$ for all comparisons). Further, $\Delta N^{Rej} = -0.08$ is equal to column 1, and significantly smaller than column 3 (1.07), column 4 (0.87) and columns 1 and 3-4 pooled ($p = 0.01$, $p = 0.01$ and $p = 0.06$, respectively).

Finally, using the sum of the ΔN as a measure of the total norm strength, the sum for Column 2 is significantly lower compared to the sum for the other three columns pooled together (1.2 vs 2.83, $p = 0.03$) and compared to the ratings of all other subjects (i.e adding in those who did not renegotiate and who completed the ratings task; 1.2 vs 2.89, $p = 0.03$). Taken together this provides support that subjects are willing to ask for renegotiation, and will not roll after a (surprise) acceptance.

The subjects in Column 3 have quite strong norms across all renegotiation outcomes, although not quite as strong as the subjects in column 1 of Table 6 (described by Hypothesis 8). In particular, $\Delta N^{Rej} = 1.07$ is largest for this

[35] We find similar results for comparing the difference to all other subjects (0.88 versus 0.27, $p < 0.01$).

group among all subjects who ask for renegotiation, with significant differences compared to Columns 1 (-0.08) and 2 (-0.08) ($p < 0.01$ for both comparisons). The strong norms associated with rolling after rejection is inline with the intuition of Hypothesis 8. However, the fact that they asked in the first place is not consistent with our simplifying assumption about subjects projecting their beliefs onto others (and hence expecting rejection). However, if these subjects instead had beliefs about acceptance closer to the empirical rate of 50%, that could justify asking. Hence these subjects are also partially consistent with our framework.

Column 4 does not provide a clean test of our framework as this group of subjects could include those described by Hypothesis 9 (who in this case were surprised by rejection), or Hypothesis 9 (who anticipated the rejection). The difference in norms across renegotiation outcomes is relatively small, which is more consistent with Hypothesis 10, however the magnitude of $\Delta N^{Rej} = 0.87$ suggests that their behavior is likely more driven by a lack of norm sensitivity (i.e. η small).

In summary, our analysis of the norm strength by sub-groups of B players shows patterns connecting the norm strength beliefs and renegotiation/rolling behavior consistent with the predictions of our social norms framework. We now turn the analysis around and use OLS regressions to predict the decision to ask for renegotiation based on patterns of beliefs in the social norm.

VIII. Predicting Renegotiation Requests With Norm Belief Patterns

For each subject we identify whether they believe the norm to be relatively strong or weak for each renegotiation outcome (i.e. after acceptance, baseline and after rejection). We create indicator variables for different configurations of ΔN^{Acc}, ΔN^{Base}, and ΔN^{Rej} for when a subject's ratings are above or below the median rating. We observe six patterns: No ratings above the median strength, baseline scenario only above, baseline and rejection ratings above, accept ratings only above, accept and baseline above, and all three above the median. Among

these patterns the two most common are believing there is no strong norm (37% of subjects) and believing that the norm is strong in all three scenarios (21% of subjects). We use these norm patterns to predict the decision to ask for renegotiation in the risky trust game.

The results are reported in Table 7. The first two columns use all the norm patterns, while the right two columns focus on the indicator for having a strong norm in all three scenarios. The odd columns use all subjects, while the even columns use subjects that chose Promise.

We find that norm strength significantly predicts the decision to request renegotiation. Subjects most likely to request renegotiation have a strong norm for acceptance but not rejection, while the least likely to request are those who have a strong norm in all three cases. Subjects that have only ΔN^{Acc} stronger than the median norm (High A) are significantly more likely to ask for renegotiation (for both the full sample and the sub-sample that promised). The joint test for High A and High AB, i.e. subjects who just have ΔN^{Acc} or have both ΔN^{Acc} and ΔN^{Base} stronger than the median norm is marginally significant ($p = 0.06$). This provides additional evidence for Hypothesis 9: Subjects who should ask for renegotiation are those whose norm beliefs that would make them optimistic about A accepting, and that would make them benefit from a successful renegotiation.

By contrast, subjects who believe that all three norms are stronger than the median (High ABR) are 40% to 60% less likely to request renegotiation, both compared to other subjects with large ΔN^{Acc} (based on coefficient differences reported in columns 1 and 2), and compared to all subjects (columns 3 and 4). This is consistent with Hypothesis 8 that subjects who believe the norm is strong after rejection will be less likely to ask, as these beliefs makes renegotiation risky, since B may end up rolling anyway but with much lower utility.

The norm data allows us to construct an estimate of the treatment effect of having the opportunity to renegotiate on final roll decisions. We nearest neighbors matching to pair each subject in the Renegotiation treatment with a subject in

Table 7—Renegotiation Requests given Norm Beliefs

Dep. Var.: B asks for Reneg.	All B (1)	If Promised (2)	All B (3)	If Promised (4)
High B	0.141	0.124		
	(0.239)	0.309		
High BR	0.316	0.183		
	(0.202)	(0.206)		
High A	0.316	0.251		
	(0.132)	(0.142)		
High AB	0.194	0.166		
	(0.236)	(0.237)		
High ABR	-0.264	-0.279	-0.413	-0.388
	(0.196)	(0.205)	(0.148)	(0.145)
Promise	0.232	0.325		
	(0.183)	(0.174)		
$X = \$7$	-0.255	-0.331	-0.236	-0.328
	(0.162)	(0.189)	(0.135)	(0.144)
Constant	0.452	0.749	0.495	0.856
	(0.162)	(0.142)	(0.159)	(0.0896)
Diff. btwn A and ABR	0.580	0.529		
	(0.177)	(0.189)		
Diff. btwn AB and ABR	0.457	0.445		
	(0.239)	(0.236)		
Observations	57	45	57	45
R-Squared	0.156	0.249	0.218	0.290

Note: Linear probability regression where the dependent variable is the renegotiation request in the risky trust game. The "High" dummies indicate whether the subject's beliefs about the norm strength are above (=1) or below (=0) the median for the situation where A accepts (High A), the baseline (High B), or above in both the baseline and the situation where A rejects (High BR) or high in all three (High ABR). Columns 1 and 3 include the ratings from all B subjects in the Renegotiation treatment who played with an A who chose In and where $X < \$10$. Columns 2 and 4 additional restrict the data to only those B subjects in the Renegotiation treatment who also promised. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

the Baseline that has the most similar norms based on ΔN^{Acc}, ΔN^{Base}, and ΔN^{Rej} (conditional on having the same Promise decision and X realization). The difference in propensity to roll between paired subjects gives the average treatment effect of having the opportunity to renegotiate.

When we look at all subjects that promised, the opportunity to renegotiate has a directionally positive effect on roll decisions ($\beta = 0.215$, s.e.=0.158, $p =$

0.174), with the effect being larger and marginally significant when $X = \$4$ ($\beta = 0.443$, s.e.=0.258, $p = 0.086$). Subjects in Renegotiation treatment that ask for renegotiation have a larger and more significant effect ($\beta = 0.339$, s.e.=0.191, $p = 0.075$), especially when $X = \$4$ ($\beta = 0.600$, s.e.=0.275, $p = 0.030$). This means that if we look at subjects who ask for renegotiation, the corresponding subjects in the control treatment with the most similar norms roll 34 percentage points less often (60 percentage points when $X = \$4$). Thus, the opportunity to renegotiate has a causal effect on roll decisions, rather than merely associating subjects that were going to roll anyways with a particular sub-sample of renegotiation outcomes. This complements our previous analysis of norm profiles which identifies which subjects will take up the opportunity to renegotiate.

IX. Discussion

We conclude our analysis by taking a step back and testing an unstated assumption that individuals care to comply with (any) norms and want to make choices that make social norms stronger. In our game this can be achieved by making a promise – which intensifies the norm. Our final analysis tests whether those B players who make a promise are different from those who do not. Table 8 reports the average norm rating for those who made a promise and those who did not in the risky trust game.

Table 8—Norm Strength by Promises

	B's choice in the behavior experiment	
	Promise	Did not Promise
$\Delta N(P$,X=\$4$) - \Delta N(NP$,X=\$4$)$	1.02	0.72
$\Delta N(P$,X=\$7$) - \Delta N(NP$,X=\$7$)$	1.12	0.70
$\Delta N(P$,X=\$10$) - \Delta N(NP$,X=\$10$)$	1.12	0.80

Note: Difference in B subjects' beliefs about the norm strength for situations where a promise was or was not made broken out by B subjects' play (did they promise or not) in the behavior section of the experiment.

B players who make a promise are those who believe that promising changes the norm to roll by a larger amount (pairwise test for $X = \$4$ and $X = \$7$, $p = 0.05$ and $p = 0.01$; test comparing the sum of differences, $p = 0.03$). This is consistent with our norms model in that following the norm gives subjects positive utility, and therefore "managing" the norms to make them stronger can be beneficial.[36]

Beliefs about the norms matter for A's In/Out decision. We focus on the norm for the baseline, as this is the only case A's are aware of when choosing to go In or Out. Those A's that chose In also believe that the norm to roll is stronger in most cases (three of the six pairwise comparisons reported in the appendix are significant, with a fourth marginally significant, and the remaining two directionally consistent). This is intuitive: If A's believe that most B's will have strong norms to roll, that increases the likelihood that A is paired with a B who will roll (and who will do so in more states or the world), which increases A's expected payoff from entering.

X. Conclusion

In this paper, we consider how promises and renegotiation outcomes impact the norm to be trustworthy. Our theoretical framework captures the intuition that promises and (un)successful renegotiations will strengthen (weaken) the norm to be trustworthy. The experimental results support this intuition: Subjects who promise are more likely to be trustworthy, especially when doing so is the most costly. Additionally, subjects' trustworthiness increases after successful renegotiations, and the opportunity to renegotiate has a causal effect on trustworthiness. The nearly 50% of subjects who do not renegotiate are more likely to believe that the norm will be strong even after a failed renegotiation. These results suggest that subjects' decisions to renegotiate and to promise are influenced by both

[36]This contrasts with other behavioral models of agreements and renegotiation (e.g. lying aversion, contracts as reference points for fairness or loss aversion, etc.), where the behavioral component creates disutility from a breach in the agreement. Absent strategic motivations subjects have little reason to magnify the psychic costs of such a break.

material considerations and a desire to manage and maximize the utility derived from complying with the norm.

Our framework differs from existing behavioral models of renegotiations and contracts in two respects. Previous models focused on providing justifications for negative reactions to renegotiation from an aggrieved party, and the decision not to renegotiate stems from a desire to avoid punishment. These models are silent in our setting, since there is no strategic consequence for B from asking for renegotiation. We present an entirely different motivation: B's choose not to renegotiate to increase the utility from being trustworthy (and hence keeping strong the norm they intend to fulfill). Second, previous behavioral models need to connect multiple behavioral forces, (as loss aversion, fairness and/or internalized social norms) to explain behavior. By contrast, we rely on one behavioral force: the desire to comply with norms.

Our contribution is to show that norms are a sufficient explanation for both when and why behaviors emerge in our setting. We show that individuals make promises, many fulfill promises as costs change, trustworthiness changes after (un)successful renegotiation, and many forego renegotiation. While our framework speaks to the trust-based transactions we consider here, the foundational assumption that a promise and/or renegotiation influences social norms by strengthening the dominant moral principle captured by the norm is applicable to any transaction (or game) where moral principles play a role. Furthermore, we provide an empirical strategy for identifying at the individual level the belief patterns about the strength of the norm that predicts specific profiles of renegotiation and trustworthiness.

REFERENCES

Akerlof, George A, and Rachel E Kranton. 2005. "Identity and the Economics of Organizations." *Journal of Economic perspectives*, 19(1): 9–32.

Andreoni, James, and B Douglas Bernheim. 2009. "Social image and the

50–50 norm: A theoretical and experimental analysis of audience effects." *Econometrica*, 77(5): 1607–1636.

Banerjee, Ritwik. 2016. "On the interpretation of bribery in a laboratory corruption game: moral frames and social norms." *Experimental Economics*, 19(1): 240–267.

Bartling, Björn, and Klaus M Schmidt. 2015. "Reference points, social norms, and fairness in contract renegotiations." *Journal of the European Economic Association*, 13(1): 98–129.

Bénabou, Roland, and Jean Tirole. 2011. "Identity, morals, and taboos: Beliefs as assets." *The Quarterly Journal of Economics*, 126(2): 805–855.

Bicchieri, Cristina. 2005. *The grammar of society: The nature and dynamics of social norms.* Cambridge University Press.

Carlton, Dennis. 1986. "The rigidity of prices." *American Economic Review*, 76(4): 637–58.

Charness, Gary, and Brit Grosskopf. 2004. "What makes cheap talk effective? Experimental evidence." *Economics Letters*, 83(3): 383–389.

Charness, Gary, and Martin Dufwenberg. 2006. "Promises and partnership." *Econometrica*, 74(6): 1579–1601.

d'Adda, Giovanna, Michalis Drouvelis, and Daniele Nosenzo. 2016. "Norm elicitation in within-subject designs: Testing for order effects." *Journal of Behavioral and Experimental Economics*, 62: 1–7.

Dewatripont, Mathias. 1988. "Commitment through renegotiation-proof contracts with third parties." *The Review of Economic Studies*, 55(3): 377–390.

Dufwenberg, Martin, Simon Gächter, and Heike Hennig-Schmidt. 2011. "The framing of games and the psychology of play." *Games and Economic Behavior*, 73(2): 459–478.

Dur, Robert, Arjan Non, and Hein Roelfsema. 2010. "Reciprocity and incentive pay in the workplace." *Journal of Economic Psychology*, 31(4): 676–686.

Eggleston, Karen, Eric A Posner, and Richard Zeckhauser. 2000. "The design and interpretation of contracts: Why complexity matters." *Nw. UL Rev.*, 95: 91.

Ellingsen, Tore, and Magnus Johannesson. 2004. "Promises, threats and fairness." *The Economic Journal*, 114(495): 397–420.

Englmaier, Florian, and Stephen Leider. 2012. "Contractual and Organizational Structure with Reciprocal Agents." *American Economic Journal: Microeconomics*, 4(2): 146–183.

Erkut, Hande, Daniele Nosenzo, and Martin Sefton. 2015. "Identifying social norms using coordination games: Spectators vs. stakeholders." *Economics Letters*, 130: 28–31.

Fehr, Ernst, and Simon Gächter. 1998. "Reciprocity and economics: The economic implications of homo reciprocans." *European economic review*, 42(3-5): 845–859.

Fehr, Ernst, and Simon Gächter. 2000. "Fairness and retaliation: The economics of reciprocity." *Journal of economic perspectives*, 14(3): 159–181.

Fehr, Ernst, Christian Zehnder, and Oliver Hart. 2009. "Contracts, reference points, and competition—behavioral effects of the fundamental transformation." *Journal of the European Economic Association*, 7(2-3): 561–572.

Fehr, Ernst, Oliver Hart, and Christian Zehnder. 2011. "Contracts as reference points—experimental evidence." *American Economic Review*, 101(2): 493–525.

Fehr, Ernst, Oliver Hart, and Christian Zehnder. 2015. "How do informal agreements and revision shape contractual reference points?" *Journal of the European Economic Association*, 13(1): 120–1.

Feltovich, Nick, and Joe Swierzbinski. 2011. "The role of strategic uncertainty in games: An experimental study of cheap talk and contracts in the Nash demand game." *European Economic Review*, 55(4): 554–574.

Gächter, Simon, Daniele Nosenzo, and Martin Sefton. 2013. "Peer effects in pro-social behavior: Social norms or social preferences?" *Journal of the European Economic Association*, 11(3): 548–573.

Gächter, Simon, Leonie Gerhards, and Daniele Nosenzo. 2017. "The importance of peers for compliance with norms of fair sharing." *European Economic Review*, 97: 72–86.

Gangadharan, Lata, Tarun Jain, Pushkar Maitra, and Joseph Vecci. 2016. "Social identity and governance: The behavioral response to female leaders." *European Economic Review*, 90: 302–325.

Halonen-Akatwijuka, Maija, and Oliver D Hart. 2013. "More is less: Why parties may deliberately write incomplete contracts." National Bureau of Economic Research.

Hart, Oliver. 1995. *Firms, Contracts, and Financial Structure.* Clarendon press.

Hart, Oliver, and John Moore. 1988. "Incomplete contracts and renegotiation." *Econometrica*, 755–785.

Hart, Oliver, and John Moore. 2008. "Contracts as reference points." *The Quarterly Journal of Economics*, 123(1): 1–48.

Herweg, Fabian, and Klaus M Schmidt. 2014. "Loss aversion and inefficient renegotiation." *The Review of Economic Studies*, 82(1): 297–332.

Iyer, Rajkamal, and Antoinette Schoar. 2015. "Ex post (in) efficient negotiation and breakdown of trade." *American Economic Review*, 105(5): 291–94.

Kessler, Judd B., and Stephen Leider. 2012. "Norms and contracting." *Management Science*, 58(1): 62–77.

Kimbrough, Erik O, and Alexander Vostroknutov. 2016. "Norms make preferences social." *Journal of the European Economic Association*, 14(3): 608–638.

Kreps, David. 1990. "Corporate culture and economic theory." In *Perspectives on Positive Political Economy.* , ed. J.E. Alt and K.A. Shepsle, 90–142. Cambridge: Cambridge University Press.

Krupka, Erin L., and Roberto A. Weber. 2013. "Identifying aocial norms using coordination games: Why soes dictator game sharing vary?" *Journal of the European Economic Association*, 11(3): 495–524.

Krupka, Erin L., Stephen Leider, and Ming Jiang. 2016. "A meeting of the minds: Informal agreements and social norms." *Management Science*, 63(6): 1708–1729.

Lambert, Nicolas, and Yoav Shoham. 2009. "Eliciting truthful answers to multiple-choice questions." 109–118, ACM.

Lyons, Bruce R. 1996. "Empirical relevance of efficient contract theory: Inter-firm contracts." *Oxford Review of Economic Policy*, 12(4): 27–52.

Macaulay, Stewart. 1963. "Non-contractual relations in business: A preliminary study." *American Sociological Review*, 55–67.

Miettinen, Topi. 2013. "Promises and conventions–an approach to pre-play agreements." *Games and Economic Behavior*, 80: 68–84.

Tversky, Amos, and Daniel Kahneman. 1979. "Prospect theory: An analysis of decision under risk." *Econometrica*, 47(2): 263–292.

Vanberg, Christoph. 2008. "Why do people keep their promises? An experimental test of two explanations." *Econometrica*, 76(6): 1467–1480.

Veselỳ, Stepán. 2015. "Elicitation of normative and fairness judgments: Do incentives matter?" *Judgment and Decision Making*, 10(2): 191.

For Online Publication: Appendix

A1. Experiment design overview

Our experiment design consists of two treatments (Baseline and Renegotiation) that collect data to test our hypotheses regarding both behavior and norms. This is a visualization of the experiment design.

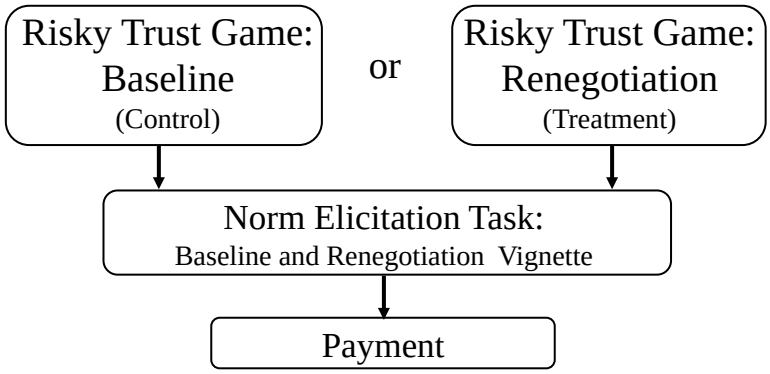


Figure A1. Overview of the Experimental Design.

Each subject participates first in either the Baseline or Renegotiation risky trust game and then in the norm elicitation task; the former collects data on choices and the latter on beliefs about the norm. The risky trust game is different depending on which treatment a subject is in but the norm elicitation task does not vary between treatments and elicits beliefs about the norms for the Baseline and Renegotiation games. Both the risky trust game and the norm elicitation task are incentivized to yield a combined final payoff.

A2. Screen shot of norm elicitation task

The figure depicts a screen shot of two of the actions subjects rated in the norms elicatation task.

FIGURE A2. PARTIAL SCREENSHOT FROM NORM ELICITATION TASK.

The screen shot provides an example of how X is varied while all other aspects are held constant. For example, after reading the baseline vignette, they are prompted with a list of possible actions available to B (two f which are depicted) and asked to play the coordination game over appropriateness ratings for each action. Subjects receive an incentive to match their appropriateness rating with that of another, randomly selected, subject (who is reading the same vignette and in the same session). They can choose between characterizing the Roll decision as anything from "very socially appropriate" to "very socially inappropriate".

A3. Additional details for individual heterogeneity

We now provide additional details for the four cases of individual heterogeneity in beliefs about the norm discussed in Section 7. As noted in the main text, since we observe an individual's elicited norm function N(), but not their norm sensitivity η, we focus on identifying characteristics of the norm function that make the set of admissible η values as large as possible.

Case 1: Not asking for Renegotiation, and Rolling

The first case we consider is where B will Roll in the base case, and also prefers not to ask for renegotiation. In the main text we argued that this should be associated with a strong norm in all three renegotiation scenarios, so that B will roll in any case and hence will expect A's to reject. This would imply $\eta\Delta N^{Rej}(X) \geq 10 - X$. Hence, when deciding to ask for renegotiation, B will compare the following two anticipated utilities:

Don't Ask, Roll: $U = X + \eta N^{Base}(R|X)$

Ask, Be Rejected, Roll: $U = X + \eta N^{Rej}(R|X)$

By Assumption 3, not asking will always be preferred. Therefore for this case to obtain it is sufficient that $\eta \geq \frac{10-X}{\Delta N^{Rej}(X)}$. This requirement is weakest (and therefore allows the most values of η) when ΔN^{Rej} is large. Given our assumptions this implies that ΔN^{Acc} and ΔN^{Base} are also large.

Case 2: Ask for Renegotiation, and Roll if accepted

The second case focused on B asking for renegotiation, and then rolling if renegotiation is accepted. In the main text we argued that this should be associated with a strong norm in the baseline and/or accepted case, and a substantial weakening of the norm if rejected. These subjects believe that A's will accept a renegotiation request, which is based on having a norm profile where B will roll if accepted, and not roll if rejected. Being willing to roll if accepted implies $\eta\Delta N^{Acc}(X) \geq (10 - X)/2$, and not rolling when rejected implies $\eta\Delta N^{Rej}(X) < 10 - X$. Given these beliefs, when considering whether to ask B will compare the following three anticipated utilities:

Ask, Be Accepted, Roll: $U = (10 + X)/2 + \eta N^{Acc}(R|X)$

Don't Ask, Roll: $U = X + \eta N^{Base}(R|X)$

Don't Ask, Don't Roll: $U = 10 + \eta N^{Base}(D|X)$

Given these subjects' beliefs, asking for renegotiation dominates simply rolling.

Hence a sufficient condition for being willing to Ask is $\eta\Delta N^{Base}(X) \geq 10 - X$ and $\eta\Delta N^{Rej}(X) < 10 - X$. This implies $\eta \in [\frac{10+X}{\Delta N^{Base}(X)}, \frac{10-X}{\Delta N^{Rej}(X)}$, which is non-empty an largest when ΔN^{Base} is large and ΔN^{Rej} is small. Alternately we could have that B will not roll in the base case. Then to have B prefer not to roll requires that $\eta \geq \frac{10-X}{2(N^{Acc}(R|X)-N^{Base}(D|X))}$. A sufficient condition is that $\eta \geq \frac{10+X}{2\Delta N^{Base}(X)}$. Therefore, this scenario can be supported if $\eta \in [\frac{10+X}{2\Delta N^{Base}(X)}, \frac{10-X}{\Delta N^{Rej}(X))}$, which contains the previous interval and again is larger when ΔN^{Base} is large and ΔN^{Rej} is small.

Case 3: Ask, Don't Roll if Accepted

The third case we discussed in the main text was associated with the scenario where Player B asks for renegotiation (expecting to be rejected), and does not roll if surprised with an acceptance. We argued that this should be associated with weak norms in all three renegotiation outcomes. If these B's anticipate rejection, then their beliefs about the norm should be such that their roll decision will be the same between acceptance and rejection (hence why they believe A has no incentive to accept). Since we are identifying subjects that won't roll if rejected (to distinguish from previous cases), that means they have to not roll if accepted as well (and therefore also not roll in the base case). To not roll if accepted implies that $\eta < \frac{10-X}{2\Delta N^{Acc}(X)}$. Given these beliefs, when considering whether to ask B will compare:

Ask, Be Rejected, Don't Roll: $U = 10 + \eta N^{Rej}(D|X)$
Don't Ask, Don't Roll: $U = 10 + \eta N^{Base}(D|X)$

This is clearly true by assumption. The major restriction is therefore $\eta < \frac{10-X}{2\Delta N^{Acc}(X)}$. which is easiest to satisfy when ΔN^{Acc} is small.

Case 4: Don't Ask, Don't Roll

The final case to consider is subjects that do not ask for renegotiation and do not roll. In the main text we argue that this should be associated with a weak baseline norm, and small increase in norm utility for being accepted relative to

the monetary benefit of not rolling. For this case B has to believe that A would accept a renegotiation request, since B intends to not roll in the baseline case. If B expected to be rejected he would prefer that than simply rolling without asking (since the disutility of not rolling would decrease). As in Case 2 this implies $\eta\Delta N^{Acc}(X) \geq (10-X)/2$, and $\eta\Delta N^{Rej}(X) < 10-X$. Additionally, since B will not roll in the base case this implies $\eta\Delta N^{Base}(X) < 10-X$. Therefore, when considering whether to ask or not B will consider these two anticipated utilities:

Ask, Be Accepted, Roll: $U = (10+X)/2 + \eta N^{Acc}(R|X)$
Don't Ask, Don't Roll: $U = 10 + \eta N^{Base}(D|X)$

To prefer not to ask requires that $(10-X)/2 > \eta(N^{Acc}(R|X) - N^{Base}(D|X))$. Therefore this case can occur if $\eta \in [\frac{10-X}{2\Delta N^{Acc}(X)}, \frac{10-X}{2(N^{Acc}(R|X)-N^{Base}(D|X))}$. This interval is non-empty, and is largest when ΔN^{Acc} is large and $(N^{Acc}(R|X) - N^{Base}(D|X))$ is small.

A4. Additional Results Analysis

Table A1—Roll Decision by Promise and X Value

Dep. Var.: B Rolling	Coefficient (1)	Marginal Effect (2)
X = 7	5.400	
	(0.517)	
X = 10	10.801	
	(0.108)	
Promise & X = 4	5.232	1.555
	(0.238)	(0.130)
Promise & X = 7	-0.211	-0.063
	(0.556)	(0.165)
Promise & X = 10	-4.092	-1.216
	(0.391)	(0.186)
Constant	-5.400	
	(0.055)	
Observations	107	
Pseudo R-Squared	0.248	

Note: Probit regression where the dependent variable is the roll decision in the risky trust game. Column 1 reports the regression coefficients, and Column 2 reports the marginal effects for making a promise. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2—Norm Beliefs by A's Trust Decision

	ΔN^{Base}					
	Promise			No Promise		
	X = 4	X = 7	X = 10	X = 4	X = 7	X = 10
A Chose Out	1.106	1.105	1.376	0.024	0.188	0.529
A Chose In	1.151	1.308	1.615	0.179	0.393	0.744
p-value	0.613	0.032	0.02	0.069	0.048	0.135

Note: Average Norm Strength for the baseline scenario for subjects in Player A role, by decision to go IN or OUT. Third row reports p-values for non-parametric rank sum tests.