Promise-keeping Norms and Renegotiation Behavior

Erin L. Krupka (School of Information, University of Michigan & IZA)[1]
Stephen Leider (Ross School of Business, University of Michigan)
Ming Jiang (Antai College of Economics and Management, Shanghai Jiao Tong University)

July 21, 2017[2]

Abstract

The desire to uphold promise-keeping norms greases the wheels of interaction by creating trust. Norms establish a set of mutual expectations which parties rely on to interact in the presence of uncertainty and renegotiation. We present a model of social norm compliance in a risky trust game. We establish a set of assumptions about the norm that characterize how promises affect the norm to fulfill an agreement, how the norm is changed once unforeseen contingencies are resolved and is changed if a renegotiation request is accepted or rejected. Using these assumptions, the model makes predictions about behavior patterns in the risky trust game. We conduct an experiment to test predictions both of the behavior patterns and the assumptions of the model. We show that behavior is consistent with the norms model and that our assumptions about the norms are supported. Using this model, we explain why most subjects make promises, why promises are largely fulfilled even when it is costly, how renegotiation success or failure affects the propensity to fulfill the promise and why nearly half of subjects do not request costless renegotiation even if it is available. This work sheds light on the impact of norms to influence renegotiation and extends the promise-keeping literature. For policies written against the backdrop of strong norms, we address implications and guidelines.

JEL Codes: C93, D23

[1]Corresponding author: ekrupka@umich.edu
[2]We wish to thank Emma Enache, Andrew Barton and Jennifer Lyu for outstanding research assistance. In addition, we wish to thank participants of the Economic Science Association conference 2016.

I. Introduction

Fulfillment of an agreement - whether formal or informal - is governed by strong proscriptive social norms. Contracts must often rely on informal agreements that are supported by norms because not all contingencies are *ex ante* knowable. "Many transactions will potentially be too costly to undertake if the participants cannot rely on efficient and equitable adaptation to those unforeseen contingencies. Note that such reliance will necessarily involve blind faith....both sides involved in a transaction must *ex ante* have some idea of the meaning of appropriate or equitable fulfillment of the contract." (Kreps 1990)[3] These *ex ante* 'ideas' are articulated in the form of principles or rules that have wide (and, as Kreps notes, preferably universal) applicability. Social norms, collectively agreed upon notions of acceptable behavior, are a likely basis for these principles of appropriate fulfillment.

This view, captured by Kreps and adopted widely, offers a vision of norms as rules of conduct that are established outside of the particulars of the transaction taking place. Their impartial legitimacy greases the wheels of trust and establishes reference points to guide behavior. In turn, parties use them to create reputation, assess the intentions of their partners, argue the merits of a request for renegotiation, and ultimately rely on them to find a way to fulfill the obligation. However, this view limits the impact that norms likely have on fulfilling agreements because it treats norms as unconditional with respect to the unforeseen contingencies and behaviors of the interacting parties. Norms can change as uncertainty is resolved and by the very presence of an opportunity for, and outcomes achieved in, renegotiation. As such, they may be able to explain much more of the behaviors that give rise to fulfilling an obligation than previously thought.

In this paper we focus on whether and how agreements are fulfilled and argue that a social norms framework can best explain the suite of behavior patterns that give rise to fulfilling an agreement. The existing literature on contracts offers a host of behavioral explanations (such as contracts acting as reference points) to explain contract fulfillment. However, Bartling and

[3]Macauley (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 et al. 2000).

Schmidt (2015) conclude that none of these behavioral theories alone is able to explain the observed behavior.

We present a norms model that characterizes how norms change in the presence of a promise, uncertainty and renegotiation. The model captures the intuition that decision makers wish to honor norms. The idea leads to a non-standard conceptualization of utility: actor's preferences depend on beliefs about the *collective* beliefs of the group regarding what one ought to do in a particular situation, i.e. the norm; utility stems from both the payoffs associated with an action and the degree to which the action complies with the norm for that situation. Using the model, we characterize how promises affect the norm to fulfill an agreement, how the norms are changed once unforeseen contingencies are resolved, and how they are changed if a renegotiation request is made. We then generate hypotheses for behavior that we are likely to observe in the risky trust game which we model.

We conduct an experiment with a stochastic trust game in the first stage and a norm elicitation task (using the Krupka-Weber (2013) method for norm elicitation) in the second stage. Using data from the risky trust game we test and show that features of behavior are consistent with the norms model. Second, using data from the norm elicitation task, we test several key assumptions of the norms model.

Our main contribution is to be able to explain a host of behaviors with one model. We are able explain why promises are made; we can explain the willingness to fulfill promises even when it is costly to do so; we are able to explain how renegotiation success or failure affects the propensity of individuals to fulfill promises, and we can explain why someone would choose to request or not request a renegotiation of an initial commitment after uncertainty is resolved. The norms framework is both flexible and powerful – making clear predictions based on easily observable features of the social norm.

A second contribution is that we shed light on the impact of norms to influence renegotiation likelihood and success. In many circumstances the shadow of potential future renegotiation affects the nature of the agreements and contracts that an individual makes. An increased understanding of how norms affect renegotiation offers important refinements to our knowledge of how to structure contracts. Many contracts that can help prevent *ex post* hold up problems are substantially

less effective if renegotiation is easy and common (for example, see Hoppe and Schmitz 2011 for experimental evidence). However, reluctance to renegotiate can lead to sticking with inefficient outcomes (see Bartling and Schmidt 2015 for experimental evidence) or, as Kreps (1990) notes, a failure to engage in trade in the first place. Iyer and Schoar (2015) describe two field experiments on hold up and renegotiation, and argue that norms of fairness, as well as reputational concerns, constrain parties' willingness to renegotiation. They give examples of the aversion to renegotiation both being beneficial (e.g. many tailors in their study were willing to rush the order without an extortionary price increase) and causing inefficiencies (other retailers refused the order outright, without seeing whether a higher price could compensate for their higher costs). Our experiment provides direct evidence for the importance of norms on renegotiation in a setting where reputational concerns are eliminated, and gives additional structure to what those norms of fairness look like.

We also provide a novel extension of the social norms and promises literature by exploring how norms for promise keeping change as a function of uncertainty and in response to actions taken by interacting parties. A parallel literature to the contracts work, shows that informal promises can have a large effect on behavior (Charness and Grosskopf, 2004; Fehr et al. 2011; Feltovich and Swierzbinski,. 2011; Miettinen, 2013; Krupka et al. 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 et al. 2011).[4] This literature also shows that making a promise elicits a promise-keeping norm which is strongly proscriptive (Krupka et al. 2016). However, no research documents the manner in which these promise-keeping norms are impacted by changes in how costly it is to honor the promise or by the actions of interacting parties. We show that the willingness to fulfill an agreement varies substantially as the context changes. Making a promise has the largest impact on behavior when fulfilling the promise is most costly. Additionally, subjects were more willing to fulfill the agreement if the other party accepted a modification of the transaction through renegotiation when large costs are realized, compared to cases where the other party rejected taking on a portion of the costs.

II. The Risky Trust Game

[4] Promises and informal agreements play a particularly important role in the context of incomplete contracts. Incomplete contracts are extremely common (Tirole 1999; Scott 2003) and can often be more efficient than other more formal contracts (Fehr and Falk 1999; Falk and Kosfeld 2006; Sliwka 2007; Rigdon 2009).

This section describes the environment we model and sets the stage for the norms model and experiment. We modify the trust game used in Charness and Dufwenberg (2006) so that we can test the norms model. The game is described so that it readily maps to the experimental design.

Figure 1 about here.

In the risky trust game, player A is able to choose whether s/he wishes to be "in" or "out" for the game, i.e. if A is willing to trust B. If A chooses "out" the game ends and both A and B receive $2. If A chooses "in", then the payoffs are determined by whether subject B chooses to "roll the die," i.e. if B is willing to be trustworthy. Our first modification to this game is adding a simultaneous decision by B of whether to make a "promise" to roll the die. Note that since this promise occurs simultaneously with A's only decision of "in"/"out", B's choice of making a promise is not conditional on knowing what A decided and vice versa. Therefore making a promise has no strategic value to B, since it cannot affect A's behavior in any way (that B knows about at this point).

The promise is a non-binding statement to "roll" in B's subsequent decision. Neither A nor B knows what X is while making the "in"/"out" and "promise" decisions, however, they both know the probability distribution for X, and that it will be determined before B's decision to roll. The value of X is determined randomly, with a 40% chance that X = $4, a 40% chance that X = $7, and a 20% chance that X = $10. By probabilistically varying the payoff associated with rolling the die, we change the cost of honoring one's promise and examine the impact of those costs on fulfilling the agreement.[5]

In stage 2, A learns whether B promised. If A has chosen "in", then B also learns what the value of X is and must make a decision about whether to "roll" or "don't roll". If B chooses not to roll then, regardless of the value of X, he or she will receive $10 and A will receive $0. After learning A's "in" decision and the realized value of X, B makes his "roll" decision, the roll outcome and payoffs are determined. Note that we follow Charness and Dufwenberg (2006) in allowing Player A to only see his monetary payoff, but not B's roll decision. If A receives a payoff

[5] The subprime mortgage crisis is one example of how the costs of keeping one's promise can change. The uncertainty around the value of X mimics interactions such as between a bank and home buyer where both parties know that there is some uncertainty about future payoffs (stemming from the realized value of the home as well as the future choices made by the buyer) if they decide to contract.

of $0, he cannot distinguish between B choosing not to roll, or B choosing to roll but getting a bad outcome.

From a standard selfish utility-maximizing 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 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".

In a second game (our treatment in the experiment) we take the basic risky trust game described above and introduce an unanticipated opportunity for renegotiation. This opportunity is presented after X is known but before B decides to "roll". In cases where X = $4 or X = $7, we give B the option of sending a message that proposes new payoffs such that A takes on (in expectation) half the cost of the negative shock.[6] A then has the opportunity to "accept" or "reject" the request and then B is faced with the "roll"/"don't roll" decision. Note that accepting renegotiation only changes the payoffs when B chooses "roll." In the standard selfish model, B is indifferent between asking for renegotiation or not, and will always choose not to roll whether A accepts or rejects. On the other hand, A is indifferent between accepting and rejecting, since it will not affect B's behavior.

The novel aspects of this trust game are that we make the "in"/"out" and "promise" decision simultaneous and we probabilistically vary the cost of honoring the promise. In our second (renegotiation) game we introduce an unanticipated opportunity for B to request, and A to accept or reject, a proposal to share the burden when X is $4 or $7. These features allow us to examine patterns of choices (eg. "promise but then don't roll" or "promise, ask for renegotiation, get rejected, then roll") for which a norms model of behavior might make unique predictions and for which norms are changed by these features. We can then use the model to generate hypotheses about behavior.

III. Model and Hypotheses

[6]If X = $4, accepting the renegotiation request means that B receives $7 if he chooses to roll, and A receives $0 with probability 1/2 and $10 with probability 1/2. If X = $7, accepting the renegotiation request means that B receives $8.50 if he chooses to roll, and A receives $0 with probability 1/3 and $10 with probability 2/3.

In this section we develop a model of decision makers that wish to honor promise-keeping norms, where the norm changes depending on the results of a renegotiation. We proceed in two steps. First, we describe a general framework that we will maintain throughout our analysis. We describe a utility function that captures norm-sensitivity, and we make six intuitive assumptions about characteristics of the norm for promise keeping. This general framework gives an explanation for behavior in the baseline treatment. Second, we extend our general framework and identify four sets of additional conditions on the norm function that explain four distinct patterns of behavior that we expect to observe in the renegotiation treatment.

Both of these steps provide experimentally testable predictions. First, the six core assumptions in our general framework are directly observable and testable from our (elicited norm) data. If concern for the social norm of promise keeping is the primary mechanism driving subject behavior, then we should observe that subjects' social norms are on average consistent with our six assumptions, and that variation in the measured social norm for each decision scenario is a significant predictor of the choices made in those scenarios. Second, the additional conditions for each pattern of behavior suggest that there should be a correlation in our subjects between exhibiting a pattern of behavior and having a particular belief about the norm. We can then compare where different subsamples of subjects with different behaviors also have different beliefs about the norm (as implied by the theory).

a. General Framework and Assumptions

In developing our general framework for social norms of promise keeping in this setting, we want our model to be able to explain the behavior in the baseline treatment that we expect most subjects to exhibit: A players choose "in", B players choose "promise", and B players are more likely to choose "roll" if they promised. Our intuition about the likely behavior of subjects in this setting is based on the existing literature on promises, e.g. Charness and Dufwenberg (2006), Kessler and Leider (2012), Krupka et al. (2016). However, while the earlier literature has primarily focused on empirically measuring the norm, we begin by identifying several intuitive assumptions about the structure that a reasonable norm of promise keeping should exhibit. We then show that these assumptions about the social norm can easily explain the expected behavior. Furthermore, these assumptions will be easily testable given our data on the elicited norm.

We assume a non-standard conceptualization of utility. We assume that actor's preferences over strategies depend on beliefs about the collective beliefs of the group, or norm, regarding what one ought to do in a particular situation. In particular, we assume that utility stems from a characteristic of the action – in this case the degree to which the action complies with the norm for that situation – rather than just the payoffs.

For simplicity we model player A as being a payoff-maximizing player and focus on player B.[7] We model player B as caring both about money and social norms because some form of non-selfish preferences is necessary to explain any player B choices that involve deciding to roll. We start with the assumption that individuals care about behaving in a manner consistent with social norms rather than developing a theory of norm compliance based on underlying preferences.[8]

We follow Krupka and Weber (2013) and Krupka et al. (2016) in defining (injunctive) social norms as collective perceptions, among members of a population, regarding the appropriateness of different behaviors.[9] To embed this definition of social norms in a simple utility framework we assume that for any action, a, a decision maker cares about both the monetary payoff produced by the selected action, $\pi(a)$, and the degree to which the action is collectively perceived as socially appropriate according to the social norm. The social norm is represented by $N(a) \in [-1, 1]$, an empirically measurable collective judgment that assigns to each action a degree of appropriateness or inappropriateness. Norm-sensitive individuals gain and lose utility for taking actions that are

[7]While player A in principle might also care about norms, his behavior can be fully explained by strategic interests, and it is less clear a priori what norms should govern A's choices. Therefore, we focus on player B. However, we could model player A's utility U_A from taking an action, "a", as depending only on his monetary payoff π_A: $U_A(a) = \pi_A(a)$.

[8]For example, Andreoni and Bernheim (2009) explore situations in which one individual unilaterally shares wealth with another, and model a norm as the behavior that results from individuals caring about own wealth, intrinsically about fairness, and about how others perceive their concern for fairness. Their paper assumes an exogenously defined alternative (x^F), on which there is implicitly agreement that it is the "fair" action for the decision maker to take, and they follow prior research in assuming that the equal (50-50) division of wealth is a natural reference point. They then show why pooling may occur at this alternative (as well as at other alternatives, under changing conditions). Andreoni and Bernheim acknowledge that x^F may differ across contexts, and may thus account for varying behavior. This is similar to our main argument, that norms change across contexts, and can therefore account for changes in behavior, and suggests that our identification of what actions people agree upon as "appropriate" or "inappropriate" might provide an empirical basis for something like x^F in their model.

[9]In this section we focus on the case that everyone has the same beliefs about the social norm N(). In the section below discussing the renegotiation treatment we will consider the possibility that different groups of subjects might have different beliefs about N() (although we will maintain that all such beliefs are consistent with Assumptions 1-6).

appropriate (N > 0) or inappropriate (N < 0). Specifically, a norm-sensitive individual puts weight $\eta > 0$ on the norm associated with an action. Player B's utility is then

$$U_B(a) = \pi(\mathrm{a}) + \eta \mathrm{N}(\mathrm{a}). \quad (1)$$

In our game, B's actions are "roll" (denoted R) or "don't roll" (denoted D), so corresponding norms will be N(R) and N(D).[10]

Additionally, we assume that the appropriateness of an action may depend on the context: specifically on the state of nature (for example, X in our trust game), on whether a promise or informal agreement has been made (denoted P) or not made (denoted N), and on whether player A has accepted or rejected a renegotiation request.

In particular, we make the following intuitive assumptions about what form the norm is likely to take. These assumptions apply to all of the cases we consider below.[11]

Assumption 1 (Norms: Promise): After making a promise, rolling is appropriate and not rolling is inappropriate for any state of nature:

$$N(R|P,X) > 0 \ \forall X \text{ and } N(D|P,X) < 0 \ \forall X.$$

Assumption 2 (Norms: No Promise): After not making a promise, rolling is more appropriate than not rolling for any state of nature:

$$N(R|N,X) > N(D|N,X) \ \forall X.$$

Assumption 3 (Norms: Rolling): Rolling is more appropriate after having made a promise than after having not made a promise for any state of nature:

$$N(R|P,X) > N(R|N,X) \ \forall X.$$

Assumption 4 (Norms: Not rolling): Not rolling is seen as less appropriate after having made a promise than after having not made a promise for any state of nature:

$$N(D|P,X) < N(D|N,X) \ \forall X.$$

[10] We focus here on norms related to fulfilment of the promise (i.e. rolling or not). Renegotiation therefore has a normative component in how it changes the norms around fulfilling the promise. In principle there may be norms directly related to the renegotiation process – i.e. whether it is appropriate or inappropriate to ask for renegotiation, or to accept or reject a renegotiation request. However, we are able to explain the decision to renegotiate without needing to appeal to normative utility directly attached to that action.

[11] Note that we will be able to test these assumptions because we elicit the social norms in the norm elicitation task in our experiment.

Together Assumptions 3 and 4 imply that for any state of nature the difference in appropriateness between rolling and not rolling is larger after promising than after not promising, i.e. $N(R|P,X) - N(D|P,X) > N(R|N,X) - N(D|N,X) \; \forall \; X$. For ease of exposition, we will denote the difference in the norm between rolling and not rolling after a promise as ΔN^{P}_{X} and after no promise as ΔN^{N}_{X}. We define norm strength as follows:

> *Definition (Norm Strength): Larger values of ΔN represent a "stronger" norm for rolling.*

What this means is that a norm-sensitive subject will have a larger incentive to choose "roll".[12] Specifically, a subject will choose to roll if $\eta * \Delta N_X \geq (10 - X)$. We should therefore expect fewer subjects to choose "roll" if X decreases.

Before introducing how renegotiation affects the norm, we first want to explore what values of η, that are consistent with our assumptions about the norm, generate patterns of behavior we might expect to see in our risky trust game. In particular, we might expect to see many A's choose "in", many B's choose "promise", and that B's who promise will choose "roll" more frequently than those that do not. In the Appendix we show that, given Assumptions 1-4, for any specific norm N() there are values of η that are consistent with these behaviors. For example, promises will make a difference in B 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. Similarly, B's will promise if the increased normative utility from rolling outweighs any decreased monetary utility from a change in behavior. Finally, A's will choose "in" if they believe sufficiently many B's will roll. Summarizing, we expect to see that A's will choose "in", B's will choose "promise" and that promises will positively affect roll decisions and that the roll decision will be impacted by the value of X. The social norms model is able to capture these behaviors and offers an explanation for why.

We now introduce two additional assumptions characterizing how renegotiation will change the norm. After an accepted renegotiation intuitively the norm to roll should be as strong, or stronger, than the base case where no renegotiation was possible - Player A was willing to lower

[12]An alternative notion of "norm strength" might be the precision and agreement in the beliefs about the norm in a population – however, we do not explore that idea in this paper.

his own payoff to make it easier for B to roll, and therefore failing to roll will be an even bigger betrayal. Conversely, a rejected renegotiation should weaken the norm, since A's unwillingness to help potentially let's B off the hook. We therefore make the following two intuitive assumptions:

> *Assumption 5 (Norms: Accepted Renegotiation): Agreeing to the renegotiation increases the appropriateness of rolling compared to the base case, while rejecting the renegotiation decreases the appropriateness of rolling for any state of nature:*
>
> $$N^{Acc}(R|X) > N^{Base}(R|X) > N^{Rej}(R|X) \quad \forall X.$$
>
> *Assumption 6 (Norms: Rejected Renegotiation): Rejecting the renegotiation increases the appropriateness of not rolling compared to the base case, while rejecting the renegotiation increases the appropriateness of not rolling for any state of nature:*
>
> $$N^{Acc}(D|X) < N^{Base}(D|X) < N^{Rej}(D|X) \quad \forall X.$$

Note that together Assumptions 5 and 6 imply that the strength of the norm for rolling increases with acceptance, and decreases with rejection relative to the baseline: $\Delta N^{Acc}{}_X > \Delta N^{Base}{}_X > \Delta N^{Rej}{}_X$ $\forall$ X. We would therefore expect that in general subjects should be more willing (than in the baseline) to roll when A accepts and/or less willing to roll when A rejects.

Taken together, Assumptions 1-6 characterize intuitive features of the social norm that we expect to be reflected by most subjects' beliefs about the prevailing social norm. In our experiment we can directly test them by examining the elicited norms from all our subjects. If valid, these assumptions can explain the most prevalent behaviors that we expect. However, to make further predictions about patterns of behavior in the renegotiation game, and to capture the more varied behavior we expect to see, we need additional conditions on (individual beliefs about) the norm function beyond Assumptions 1-6.

b. Additional Conditions on Beliefs About the Norm Function

When renegotiations are possible, we expect to see a more complicated set of behaviors. First, we expect many subjects to have "successful" renegotiations: B asks for a renegotiation, A accepts, and then B fulfills the (implicit) agreement by rolling. Additionally, many player B's may choose not to renegotiate. We identify two reasons why a B may prefer not to renegotiate. In one case, B may be willing to roll even if A rejects renegotiation. For this reason, B may prefer not to request renegotiation because B would avoid risking the lower norm utility for rolling after a rejection

(compared to the base case). In the other case B's may prefer not to roll; for this reason, he chooses not to ask so as to avoid being accepted and then feeling obligated to roll. Finally, a B may ask for renegotiation even if they know they will be rejected in order to benefit from the reduced disutility from not rolling after a rejection (compared to the base case).

Why might one expect more varied behavior in the renegotiation treatment? Note that there are a variety of specific individual beliefs about the norm (e.g. whether ΔN is large or small) that are consistent with Assumptions 1-6, and depending on these additional details different patterns of behavior are possible. For example, Assumptions 5 and 6 only restrict ΔN^{Rej} to be less than ΔN^{Base}; it could be that ΔN^{Rej} is large or small in magnitude. If B expects a renegotiation request to be rejected, then the magnitude of ΔN^{Rej} will determine whether requesting renegotiation is an attractive option for B or not.

We anticipate that the different patterns of behavior can be explained by these individual differences in beliefs about the norm function (while still constraining ourselves to the set of norm functions consistent with our six general assumptions). In order to explain these patterns of behavior we identify several additional conditions on (beliefs about) the norm function that would support an equilibrium that would generate each pattern of behavior. These conditions provide additional testable predictions in the renegotiation treatment. The elicited norms among subjects exhibiting each pattern of behavior should differ in ways consistent with the specific conditions on the norm beliefs for each behavioral pattern.

We note here how we are thinking about individual heterogeneity. Individuals could differ in at least two respects: an individual's norm sensitivity parameter η, or in their beliefs about the social norm N(). In our experiment we will not be able to directly measure an individual's η, but do observe their elicited belief about the norm. We can therefore compare the elicited norms of subjects that exhibit different behaviors. To that end, we identify observable characteristics of N() that affect the size of the set of η values consistent with that behavior. When the set of possible η values is larger, than it should be more likely for an individual to exhibit that behavior. Similarly, we should expect most subjects exhibiting a particular behavior to have beliefs about the norm consistent with making the set of η values large.

We now describe in more detail the conditions on the norm associated with each pattern of behavior. Further details are in the Appendix; here we focus on the intuition, and note in bold the additional conditions on N() for each pattern. In each case we also state the expected correlation between behavior in the risk trust game and beliefs about the norm.

Successful Renegotiation: Successful renegotiation involves Player A accepting the renegotiation request. This only occurs if accepting increases the likelihood of B choosing "roll". Given that A will accept renegotiation, a B player is willing to ask for renegotiation if (a) he was going to roll in the base case, or (b) he was going to not roll in the base case, but the utility benefit of rolling if accepted is large enough to induce him to ask. In the Appendix we identify a non-empty interval of η values that are sufficient for this condition to obtain. The interval is largest (and hence easiest to satisfy) when $\mathbf{\Delta N^{Base}}$ **is large and** $\mathbf{\Delta N^{Rej}}$ **is small**. Note that as a consequence of Assumption 5 and 6, ΔN^{Base} being large means we should also see **large** $\mathbf{\Delta N^{Acc}}$. We therefore expect most subjects that have successful renegotiations to also have beliefs about the norm involving a large ΔN^{Base} and ΔN^{Acc}, and a small ΔN^{Rej}.

Not Asking for Renegotiation and Rolling: Player B's may prefer not to ask for renegotiation if they anticipate being rejected and rolling anyways, since rolling after renegotiation leads to lower norm utility than rolling in the base case. This belief is consistent, since if Player A anticipates that B would still roll if rejected then he will in fact reject the request. This scenario is most likely to occur if η is large and/or **all** $\mathbf{\Delta N}$ **(but most importantly** $\mathbf{\Delta N^{Rej}}$**) is large**. For these highly norm-sensitive subjects, they prefer simply to roll rather than risk the lower norm utility for rolling after rejection. Hence, we expect subjects that do not ask for renegotiation but do roll to have beliefs about the norm such that ΔN^{Rej} is large.

Not Asking for Renegotiation and Not Rolling: In this scenario B chooses not to ask because he anticipates being accepted, but prefers the higher monetary payoff from not rolling to the increased normative payoff of being accepted and rolling. This is more likely to occur when the normative payoff only increases by a small amount, i.e. if $N^{Acc}(R|X) - N^{Base}(D|X)$ is relatively small. We would therefore expect the beliefs about the norm of subjects that do not ask for renegotiation and do not roll to imply that $\mathbf{N^{Acc}(R|X) - N^{Base}(D|X)}$ **is small**. Additionally, since this groups of subjects would roll if they asked for renegotiation and were accepted, we should

still see a fairly **large value for ΔN^{Acc}**. Similarly, since they are not rolling in the base case we should see **ΔN^{Base} small.**

Asking for Renegotiation Expecting Rejection: Finally, a Player B may want to ask for renegotiation even if he knows (believes) he will be rejected. If B knows he will not roll in any case, then asking and being rejected reduces his disutility for violating the norm. This scenario can occur when η is small and/or **ΔN is small (in particular we need ΔN^{Acc} to be small)**. Additionally, given Assumption 6, in this scenario it is not just the case that B expects to be rejected – he actually receives a higher utility by being rejected than he would receive if accepted. One challenge for testing this pattern of behavior is that we don't directly observe B's expectations when asking for renegotiation, and looking at all subjects that had a rejected request may include subjects that were expecting a successful renegotiation but were surprised by a rejection (who would also choose not to roll after a rejection). We will therefore look for this correlation among subjects that had an accepted renegotiation but chose not to roll anyways – i.e. subjects from this scenario that were surprised by a rejection.

c. Alternate Behavioral Models of Renegotiation

These scenarios show that the social norms model is both flexible enough to capture various different patterns of behavior that we expect to see in our data, while also providing testable predictions about what else we observe about the different subjects in each case. Two other prominent behavioral models of contracts (Hart and Moore, 2008 and Herweg and Schmidt, 2014) have been used to explain renegotiation behavior in other experimental settings (Fehr et al., 2011; and Bartling and Schmidt, 2015). In the Hart and Moore model the initial agreement sets a reference point for what fair treatment consists of under the agreement. If one party uses discretion or renegotiation to give the other party a worse outcome than he was expecting under the initial agreement, this party may feel aggrieved and use their *ex-post* actions to punish the first party. In the Herweg and Schmidt model renegotiated outcomes may be interpreted as losses when they are worse for one party than the initial agreement would indicate.

How well can these other behavioral models of renegotiation apply to our setting? First, note that we do not have a formal contract in our setting, so the most direct interpretations of these models would simply not apply in our context. If we treat the promise as the contract, we still

cannot account for why a B player would not request a renegotiation using this type of contract-as-reference point model. In Hart and Moore (2008), a party compares an *ex-post* outcome to the most favorable outcome permitted by the contract. In our setting Player A is the only one made worse off by the renegotiation – and therefore should be the only one that feels aggrieved. However, unlike in the experimental setting in Fehr et al. (2011), where the aggrieved party (the supplier) takes another action (determines quality) where they can punish the offending party, in our setting A doesn't take any other action after the renegotiation. The most A can do is reject the renegotiation. From B's perspective, there is no downside to asking and getting rejected and there is nothing in the "contracts as reference point" explanation that suggests that B should interpret a rejection as an aggrievement, or avoid renegotiation in anticipation of such an aggrievement. To the extent that B's reluctance to roll after a rejection is coming from feeling aggrieved it cannot come from the "contracts as reference points" mechanism, because the renegotiation was a surprise in our design and therefore the initial reference point agreement would not have assumed a successful renegotiation.

Turning to an explanation that relies on "loss aversion relative to the initial contract" (Herweg and Schmidt, 2014), we again find it difficult to predict why so many B's would fail to request a renegotiation. In this explanation, the reference point is the original contract given the realized state of the world. In our context, B has no losses. This is because if A rejects, then B simply does what the original agreement was. In short, the renegotiation would translate to a loss for A but not for B, and therefore loss aversion does not explain why B does not ask.

Finally, even in the most generous interpretation, both models would place the primary difference in behavior between rejection (leading to aggrievement or loss) relative to acceptance or the baseline. In both models the only reason why acceptance might lead to more rolling is from the change in the material payoffs. By contrast, the social norms model can explain an increase in rolling after acceptance separate from the material consequences because ΔN^{Acc} is larger than ΔN^{Base}.

Outcome based fairness models, such as inequality aversion (Fehr and Schmidt 1999) have also sometimes been applied to renegotiation settings. However, such models have little explanatory power in our setting. First, since these models depend only comparing material

outcomes, and a promise has no impact on the material payoffs, there should be no difference in B's behavior based on whether B made a promise. Second, given the payoffs we use, the core assumptions (of behindness aversion being weakly larger than aheadness aversion and aheadness aversion being less than 1) imply that B would always weakly prefer to ask for renegotiation. Hence models of this sort cannot explain why a substantial number of B's might choose not to ask for renegotiation. Finally, to explain roll decisions by B requires implausibly large aheadness aversion values. For typical parameter values[13], inequity averse subjects would never roll in the baseline, nor would they roll when X = \$4 even if A accepts renegotiation in the treatment.

In summary, two important behaviors in the renegotiation treatment cannot be explained by the reference-point based models: that B's will be more likely to roll after an acceptance than a rejection, and that many B's will choose not to renegotiate. Similarly, outcome-based fairness models cannot explain the impact of promises, B's willingness to roll, or B's choice not to ask for renegotiation. We have shown how all of these outcomes can be explained within a social norms framework given appropriate assumptions about the shape of the norm, but that other behavioral models typically applied to renegotiation cannot. With these assumptions about the norm and hypotheses about behavioral patterns in hand, we now turn to our experiment to test for their presence.

IV. Experimental Design

We need an experimental design that will collect data on both behavior and norms in order to test our predictions. Our experiment consists of a *within* subject design with two different tasks as well as a *between* subjects design for the treatments. Figure 2 depicts the flow of the experiment and gives an overview.

Figure 2 about here.

Each subject participates first in the risky trust game and then in the norm elicitation task. The risky trust game always precedes the norm elicitation task. However, subjects either play a baseline

[13] We consider four common values for behindess aversion (α) and aheadness aversion (β): (α, β) = (0.5, 0.25), (1.0, 0.6), (2.0, 0.6) and (4.0, 0.6). For B to be willing to roll, one generally needs to assume β sufficiently close to 1 and α sufficiently close to β. For these parameter values, B always prefers "don't roll" for X = \$4 and \$7 in the baseline, and for X = \$4 after acceptance in the treatment condition.

version of the risky trust game (with no renegotiation option) or a renegotiation version (with a surprise renegotiation in stage 3). Both the risky trust game and the norm elicitation task are incentivized to yield a combined final payoff.

a. Risky Trust Game Experimental Design

Table 1 depicts a summary of choices and associated payoffs for the baseline risky trust game. These choices and payoffs are common knowledge as subjects make their "in"/"out" and "promise" decisions in stage one.

Table 1 about here.

The renegotiation treatment is identical to the baseline through stage 2. However, in stage 3, if X is equal to either \$4 or \$7, then she is informed that she has a chance to renegotiate the payoffs before she makes the "roll" decision. For example, if X = \$4, then B is told,

> "*You now have the option of asking player A if s/he is willing to change the dice game. If you choose to do so, the following message will be sent to player A: "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. The following tables summarize the payoffs for this proposal."* [If X=7, then the text is identical but we modify the proposal such that B gets \$8.50 (column 4 of table 2)].

Table 2 depicts the payoff table in the renegotiation treatment. Subject B can decide to send a message and request renegotiated terms (this step is depicted in Figure 2 at time T = 2 in stage 3). If B does so, then player A is informed of the request and the new proposal. A has the choice to "accept" or "reject" the renegotiation proposal by either sending a message to B that states "No, I don't want to change our agreement." Or "Yes, I want to change our agreement" (this is depicted in Figure 2 at time T= 3 in stage 3). B receives the message and then proceeds to make his "roll" decision. The game ends after this and subjects receive instructions about the second task.

Table 2 about here.

b. Norm Elicitation Task Experimental Design

To measure the extent to which actions are jointly recognized to be socially appropriate or inappropriate we follow the Krupka and Weber (2013) norm elicitation protocol which involves having subjects play coordination games over rating how socially appropriate it is to take actions in a scenario described to the subjects. Stage 6 is the norm elicitation task. This task proceeds in an identical manner for subjects in the baseline and renegotiation treatment of the risky trust game.

In this task, subjects read vignettes that each describe the baseline or renegotiation risky trust game respectively. The vignettes focus on B's choices. After reading the vignette, subjects are asked to evaluate different possible actions available to B and to decide, for each of the possible actions, whether taking that action would be "socially appropriate" and "consistent with what most people expect individual B ought to do" or "socially inappropriate" and "inconsistent with what most people expect individual B ought to do." Subjects are incentivized to give appropriateness ratings for the actions that match those of the other subjects in the session with them. They are paid based on the similarity between their appropriateness rating and that of one other participant who is randomly selected. We use a proper scoring rule described below.

The *baseline vignette* describes the baseline risky trust game. Once described, we list all possible combinations of the "roll" and "promise" actions when X is $4 and $7. For example, here is an excerpt of what subjects read in the *baseline vignette*:[14]

> *In the first stage of the game A is able to choose whether he or she wishes to be IN or OUT for the game. If A chooses OUT the game ends and both A and B receive $2. If A chooses IN the payoffs will be determined by whether player B chooses to roll the die.*
>
> *While A is making this decision, B must decide whether to make a non-binding promise to roll the die. If A chose IN, the payoffs are determined by B's actions. If B chooses to NOT ROLL, he or she will receive $10 and A will receive $0. If B chooses to ROLL the payoffs are as follows: If the die comes up 1, then A receives $0 and B receives $X; if the die comes up 2,3,4,5 or 6, then A receives $10 and B receives $X. The amount of money that X is equal to will be determined by the computer, with a 40% chance that X=$4, 40% chance that X=$7, and 20% chance that X=$10.*

[14] This text is abbreviated and a full version is available in the appendix.

The table on the following screen presents a list of possible actions available to Individual B. For each action, please indicate the appropriateness rating that someone selected at random would provide for each action.

The table then shows all of the possible actions that B could take *conditional* on whether she did or did not make a promise, the value of X and whether B rolled.[15] Figure 3 depicts two rows from the table subjects used to enter their responses.

Figure 3 about here.

Figure 3's two rows and two columns highlight the ways in which we vary the details of the *baseline vignette*. In row one in the left column, B's "not roll" action is taken in the context of having made a promise and X being $4. However, row one of the right column describes the same situation as the left column (B made a promise and X = $4) except that now B has decided to take the action "roll". The left and right columns of row 2 have a parallel process of systematic variation in whether B has promised except that now X = $7. The full screen shot is depicted in the instructions appendix, however, we can summarize that the actions in the table vary the "promise", "X-value" and the "roll" aspects of the vignette to create 8 unique combinations of the "roll" decision.

For each row the subject plays a coordination game with another subject over appropriateness ratings. 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". Thus, respondents play a coordination game in which the incentive is to anticipate the extent to which others will rate an action as socially appropriate or inappropriate, and to respond accordingly.

Krupka and Weber (2013) show that collectively-recognized social norms create focal points in the matching game.[16] If there is a social norm that some actions are more or less socially appropriate,

[15]The decision screen is depicted in the Appendix.

[16]Camerer and Fehr (2004) note that coordination games can be used with economic incentives to reveal shared understanding. For others using the Krupka and Weber protocol, see also Burks and Krupka (2012); Krupka et al. (2016); Kimbrough and Vostroknutov 2014, Gaechter et al. 2013; Vesely 2015; Erkut et al. 2014; D'Adda et al. 2015.

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 measure of the social norm.

After completing these ratings, subjects read the *renegotiation vignette.* This vignette describe the risky trust game with renegotiation. Similar to the *baseline vignette*, we list all possible combinations of the "roll" and "promise" actions when X is \$4 and \$7 but now we also vary whether A has "accepted" or "rejected" a request from B for renegotiation. For example, here is an excerpt of what subjects read in the *renegotiation vignette*:

> *If X is not equal to \$10, B is allowed to send a message to A asking to change the agreement if B chooses ROLL. Neither A nor B knows this option exists until the value of X is determined. If X=\$4, then B can ask that the payoffs be changed so that A gets \$0 and B gets \$7 if the die roll comes up 1,2, or 3 and A gets \$10 and B gets \$7 if the die roll comes up 4,5, or 6. A can accept B's offer or elect to keep the original payoffs. Once the payoffs have been decided, B chooses whether to ROLL or NOT ROLL….In all of the following scenarios X=\$4.*

Similar to the table excerpted in figure 3, subjects are shown a table that has all of the possible actions that B could take *conditional* on whether she did or did not make a promise, whether A accepted the renegotiation offer and whether B rolled.[17] This entire sequence is repeated for a situation where X = \$4 and X = \$7.

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[18] about the distribution of others' ratings by matching a subject with another subject and then paying them according to the following payoff function:

$$\pi_i = \$15 - \$4|x_i - x_{-i}|, \text{ for each subject } i \qquad (2)$$

[17]The decision screen is depicted in the Appendix I.

[18]We chose to elicit an estimate of the median because this yields fewer extreme ratings when the distribution of the other's ratings is particularly skewed (as might be the case for actions that are, as an example, extremely self-regarding or other-regarding). Further, while there may be no changes in the modal rating an action receives, the median rating can change between treatments. As an example, even if the modal rating for taking the most pro-social action is unchanged when there is an agreement or not, the degree to which appropriateness ratings vary for actions that deviate from the most pro-social action may vary when an agreement is in place. This, in turn, will change the median rating.

where π_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.[19] In short, this scoring rule pays them \$10 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.

In order to test our behavior hypotheses and norms assumptions, we converted subjects' norm ratings into equally spaced numerical scores. A rating of "very socially inappropriate" received 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.

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.

V. Results: Behavior in the Risky Trust Game

We now begin by discussing the basic behavior we observed in our risky trust game. Recall that based on our general framework for social norms of promise keeping (presented in section III), we expect most subjects to make the following choices: A players choose "in", B players choose "promise", and B players are more likely to choose "roll" if they promised. In analyzing our data we first test whether we observe these behaviors in our experiment.

[19]The formal proof of how this payoff function elicits a rater's guess about the median response can be found in Lambert and Shoham's 2009 paper. For our subjects, this payoff function pays them \$15 if they match the other person's ratings exactly. For each category by which they differ from their matched counter-part, subjects lose \$4. Thus, if they are off by 1 category in either direction, then they are paid \$11 and so on. If they are off by five categories (the most they can be wrong) then they pay the experimenter \$5 (ie, in the worst case they would lose their show-up fee). The intuition is that the symmetry of the penalty is sufficient to eliminate any bias in guesses since the rater has an equal incentive neither to be above nor below the median rating. Second, by making the penalty proportional to the difference in one's own rating and that of the other rater, we properly incent guessing about the median. Taken together, the symmetry of the penalty and an increase in penalty that is proportional to the degree of error, make this a proper scoring rule for eliciting a subject's guess about the median response. Finally, we choose a relatively large penalty for miss-coordination (\$4 per category difference) in order to reduce the potential for bias coming from risk aversion, where individuals could bias their ratings towards the middle rating in order to reduce the variance in their coordination payoffs.

The baseline and renegotiation treatment do not differ until after A makes his "in"/"out" decision and B makes his "promise" decision; for this reason we do not expect differences between the baseline and renegotiation treatment for A and B's first decision and we pool the data for the analysis of this first decision. We find that overall 76% of A's decide to go "in" and that this difference is not significant by treatment (72% in the baseline and 78% in the renegotiation treatment; proportions test diff. n.s.). We also find that overall 82% of B's "promise" and that this difference is not significant by treatment (83% in the baseline and 82% in the renegotiation treatment; proportions test diff. n.s.).

Figure 4 about here.

Restricting our analysis to the baseline treatment, we also find that making a promise affects the "roll" decision. Figure 4 shows the "roll" decision as a function of whether a promise was made and the value of X as either \$4, \$7 or \$10. For when there both is and there is not a promise, the probability of rolling decreases with lower values of X. In particular, rolling when X = \$4 is 32%, when X = \$7 is 43% and when X = \$10 is 92% (non-parametric trend test; $p < 0.01$). However, promises 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) verifies the effect of "promise" depends on X as well (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, $p<0.01$).

Our results for the baseline treatment are well in line with the behavior we expected given Assumptions 1 to 4: most subjects choose "in" and "promise", and making a promise significantly affects B's decision to roll. We now turn to results from the renegotiation treatment. We begin by giving a summary of the behaviors we observe and then highlight the effects (that are directly implied by Assumptions 5 and 6) of accepted and rejected renegotiation on roll decisions.

Figure 5 about here.

Figure 5 depicts the fraction of B's who request a renegotiation by whether or not they made a promise and by the value of X (either \$4 or \$7). Only about half of B subjects ask for a renegotiation (56%) and those requests increase when X = \$4 (69% vs. 40%; proportions test; $p <$

0.01). However, the impact of X = $4 on requests is entirely driven by those B's who have made a promise (79% vs 38%; proportions test; $p < 0.01$; diff. n.s. for the "no promise" comparison). Further, requests for renegotiations are higher when X = $4 and there is a promise (79%) than when X = $4 without a promise (38%; proportions test; $p < 0.01$). There is no significant difference between the fraction of requests when there is or is not a promise when X = $7. A regression (reported in the appendix) verifies what the nonparametric tests report: the effect of the promise on requesting renegotiations depends on the value of X (the marginal effect of making a promise on the decision to renegotiate for X = $4, and X = $7: +0.387, $p<0.05$; -0.1.03, diff. n.s.).

In section III, using Assumptions 5 and 6, we identified two behaviors that we would expect to see in the renegotiation treatment. In particular, we expect to see that more B's are willing to roll if the renegotiation request is accepted and fewer B's roll if it is rejected. About half (47%) of A subjects accept the request for renegotiation and there is no difference in accepting the request when X = $4 or $7 (about 45% when X = $4 and 50% when X = $7; proportions test, diff. n.s.). Further, the fraction of A subjects who accept the request does not differ by whether B has or has not made a promise (about 40% grant the request when there is no promise and about 48% when there is promise; proportions test, diff. n.s.). However, the regression results (reported in the appendix) show that the effect of a promise on accepting a request depends on X such that A's are more likely to accept the request when a promise is made and X = $7 (the marginal effect of a promise on accepting renegotiation for X = $4, and X = $7: -0.227, diff. n.s.; +1.934, $p<0.01$).

Figure 6 about here.

As hypothesized by Assumption 5 and 6, A's decision to accept or reject the renegotiation request significantly impacts B's subsequent "roll" decision. Figure 6 depicts the percent of B's who decide to roll conditioned on whether a renegotiation was rejected, B had the option to renegotiate but did not ask, or B asked and the request was accepted. Accepting renegotiation significantly increases the chance that B rolls - 67% roll if the request is accepted while only 35% roll if rejected and only 24% roll if they did not ask to renegotiate (proportions test comparing the roll decision if the request was accepted vs rejected has $p=0.08$, while a comparison of "accept" vs "did not ask" is $p < 0.01$, and "accept" vs pooling the rejections and those who "did not ask" together is $p < 0.01$). Regression verifies this result when controlling for X and "promise"; the

marginal effect of "accept" is +0.365, p<0.01; of "reject" is +0.062, diff. n.s.; the difference between "accept" and "reject" is significant at p<0.05 where the omitted group is those who did not ask).[20,21]

Table 3 about here

Table 3 summarizes the share of subjects following each path of play when renegotiation is possible (i.e. conditional on A choosing "in" and X being $4 or $7), for all subjects and for B subjects that promised. The last two columns notes if the path of play is consistent with an equilibrium (either as the on-equilibrium path, or the path if A deviates). Of the approximately forty percent of subjects that do not request renegotiation, a quarter (one third if B promised) choose to Roll. The two most common paths after a renegotiation request is that A accepts and B rolls, or A rejects and B doesn't roll (both representing approximately one third of the cases with renegotiation). Finally, in approximately one sixth of renegotiations A accepts and B doesn't roll, or A rejects and B rolls anyways.

VI. Results: The Norm Elicitation Task

Having shown that features of observed data in the risky trust game are consistent with the norms model, we now turn to testing the six assumptions about the norms which we outlined in our general framework. The six core assumptions are directly observable and testable from our elicited norm data. Here we demonstrate that subjects' social norms are on average consistent with

[20]The results are similar if we include all the data (i.e. adding in subjects where X = $10, as well as the data from our baseline treatment). The marginal effect of "accept" = +0.343, p<0.01, "reject" = +0.047, "did not ask" = +0.028. The difference between "accept" and "reject" is significant at p<0.05, between "accept" and "did not ask" is significant at p<0.05 where the baseline treatment is our control.

[21]We can also compare the average profits that players receive, based on the decisions they make during the risky trust game. A players earn significantly higher profits by choosing "in"; they earn $4.39 when choosing "in" and $2.00 when choosing "out" (pooled baseline and treatment OLS; beta = 2.39, p<0.01); the same result is obtained when just using data from the baseline; $4.41 vs $2.00; OLS Beta = 2.41, p<0.01). However, the B's earn directionally lower profits by promising ($7.16 vs. $8.04; pooled baseline and treatment OLS Beta = -0.88). When we partition our data so that we just look at the baseline, the profits are directionally higher when promising ($7.33 vs $6.25; OLS Beta = 2.41). However, they are significantly lower when promising in the renegotiation treatment ($7.07 vs. $8.88; OLS Beta = -1.81, p<0.05). B's earn marginally significantly lower profits by requesting renegotiation ($8.27 vs $9.16; OLS Beta = -0.89, p<0.10) and A's earn directionally more by accepting the renegotiation request ($4.67 vs. $3.53; OLS Beta = +1.14).

the six assumptions, and that variation in the measured social norm for each vignette is a significant predictor of the choices subjects made in the risky trust game.

To test these assumptions, we use the data collected from the norm elicitation task. Recall that regardless of their role in the risky trust game (whether they were an A or a B player) and regardless of the treatment they were in (whether they were in the baseline or renegotiation treatment), all subjects read the same vignettes in the norm elicitation task and played the coordination game. The responses from the norm elicitation task are used to generate the norms data used in this analysis. Unless otherwise specified, we are pooling all responses across treatments and player roles in the risky trust game. We use the average norm rating as our empirical estimate for the norm for a particular action unless otherwise specified.

Figure 7 about here.

Assumption 1 (Norms: Promise) and Assumption 2 (Norms: No Promise) characterize how norms differ for the action of "roll" and "don't roll" for all values of X. Figure 7 depicts the elicited average norm ratings in the *baseline vignette* for those actions. In the left panel, where B has not made a promise, the average ratings for "roll" are always higher than for "don't roll"; put differently, when B has not made a promise "roll" is more appropriate than "don't roll" and this is consistent with Assumption 2. In the right panel, where B has made a promise, "roll" is considered very appropriate (above the 0-axis) and "don't roll" is inappropriate (below the 0-axis) and this is consistent with Assumption 1. Figure 7 also shows that the data are consistent with Assumption 3 (Norms: Rolling), that "roll" is more appropriate after having made a promise than after having made no promise and that "don't roll" is less appropriate after having made a promise than after having not made a promise (Assumption 4 (Norms: Not rolling)). These differences in the norms are significant for both the "promise" and "no promise" states (signed rank test $p<0.01$ for both).

Figures 8 and 9 about here.

We now turn to the *renegotiation vignette* to test how the norm changes after an accepted or rejected renegotiation; we compare ratings to those from the *baseline vignette* without renegotiation. Figures 8 and 9 depict the mean norm ratings for the "roll" and "don't roll" action

where X is either \$4 or \$7. Figure 8 depicts the case where B has made a promise and Figure 9 the case where B has not made a promise.

Assumptions 5 and 6 together imply that the following inequalities should hold about the norms ratings in the renegotiation treatment: the change in appropriateness of the norm for rolling increases with acceptance, and decreases with rejection relative to the baseline:

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

Table 4 reports the average norm strength (ΔN) for each scenario. Consistent with the assumptions, we find that in all cases an accepted renegotiation increases the strength of the norm to roll, while a rejection decreases the strength of the norm (signed rank test $p < 0.01$ for all).

Table 4 about here.

So far, we have shown pairwise scenario comparisons of subject behavior that are consistent with the norms model, and shown that the norms themselves exhibit the structural characteristics that we outlined in Assumptions 1-6. Now we more directly show the connection between norms and behavior by using a unified conditional logit analysis to demonstrate that across all the contexts that we consider, the variation in behavior is explained by variation in norms.

In order to test for the correlation between norms and behavior we assume that individuals have a logistic choice rule, where the likelihood of choosing any action, a, depends on the relative utility of that action compared to the other action, with the utility U_i depending on the material payoff and the social norm as described in equation (1). We write this choice probability as,

$$\boldsymbol{P(a = a_i) = \frac{\exp(U_i)}{\sum_j \exp(U_j)}}. \quad (3)$$

To test whether concern with norm compliance guides behavior, we can estimate weights on the two components of equation (1) using the average appropriateness ratings which were obtained in the norm elicitation task and the choices made in the risky trust game. In the conditional logit regression (McFadden 1974)[22], the dependent variable is the "roll" decision and the independent

[22] Conditional logit models are similar to multinomial logit models, however conditional logit models emphasize the characteristics of the alternatives, while multinomial logit models depend on the characteristics of the individual

variables are the characteristics of the possible action choices - specifically each action's social appropriateness and its expected monetary payoff. The variation used for identification stems from variation across the characteristics of the possible actions. For every realization of X (and so for every set of monetary payoffs) we observe choices in both the "no promise" and the "promise case" – hence we observe choices in situations with identical monetary payoffs but differing norms. Thus, differences in behavior in the presence or absence of a promise have to come from the norm component of utility, and identify the relative importance of norm compliance in subjects' utility function.

Table 5 reports the conditional logistic regressions where we test the impact of norm ratings on the "roll" decision. Because we elicit norms using a within subject design (where subjects in both the baseline and renegotiation treatment of the risky trust game provide norm ratings in the norm elicitation task), we run the regression two ways. Column (1) reports the regression using the average elicited social appropriateness rating as a proxy for the social norm N(a), and column (2) uses each subject's individual guess about the norm as a proxy for the norm N(a). The coefficient on "norm of action"[23] provides an estimate of the weight on social appropriates, η, in equation (1) such that,

$$\boldsymbol{u_i}(\boldsymbol{x}, \mathbf{a}) = \boldsymbol{\beta x}(\boldsymbol{a_i}, \boldsymbol{a^{-i}}) + \boldsymbol{\eta N}(\boldsymbol{a}). \quad (4)$$

In each regression, the reported coefficient reflects the relative weight that each component has in the utility function. In column (1) the coefficient on "monetary payoff", though positive and significant (β=0.379; $p<0.01$), is about a third of the magnitude of the coefficient on "norm of action". The coefficient on "norm of action" is positive and significant (β=1.086; $p<0.01$). This indicates that subjects put substantial weight on following the social norm, and are willing to give up monetary value to take actions that are deemed more socially appropriate. We see a similar pattern if we use the individual norm ratings (column 2) rather than the average norm ratings to predict "roll" decisions. The coefficient on "monetary payoff" is positive and significant (β=0.352;

making the choice. See Hoffman and Duncan (1988) for a comparison between these models. Our econometric approach follows Krupka and Weber (2013) and Leider et al. (2016).

[23]We restrict η to be the same for everyone.

p<0.01) but the coefficient on "norm of action" is larger, significant and positive (β=0.824; p<0.01).[24]

Table 5 about here.

Having tested that Assumptions 1-6 (which characterize for this setting intuitive features of the social norm that we expect the vast majority of subjects to exhibit), we now test several additional assumptions about the norm function that we used to predict additional patterns of behavior.[25] The theory predicts that there will be a correlation between exhibiting one of those patterns of behavior (observed in the risky trust game) and having a particular belief about the norm (observed in the norm elicitation task). To test these predictions about the norm we examine whether different sub-samples of B subjects, identified based on their behavior in the risky trust game, also have different beliefs about the norm. For each risky trust game behavior pattern, the theory indicates that certain aspects of the norm should exhibit large or small differences. For example, each pattern requires that the norm for rolling in certain cases must either be strong (i.e. ΔN large) or weak (i.e. ΔN small). We show that for each pattern the relevant sub-sample of subjects also has significantly different norms (relative to other sub-groups) in accordance with our theory.

Table 6 about here.

Table 6 reports the norms strength for three groups of B players in the risky trust game divided by B's actions of asking for renegotiation and/or rolling. We focus here on behavioral sequences that depend only on B's choices – in Table 7 we will break things apart further based on A's response to the renegotiation request. The reported norms are for the relevant decision context (X value and promise) that the subject faced in the risky trust game.

[24]We have (weak) evidence that the average norm gives more explanatory power. The average norm ratings have slightly better LL, AIC and BIC scores however, directly testing the two models is not significant by a Vuong test (p = 0.15, one-tailed), but is significant by a Clarke's Distribution-Free test (p = 0.06, one-tailed). Perhaps more interesting is that if we run a specification with both the average and the individual norm ratings together in the model, then the coefficient on the average norm ratings is marginally significant and has a larger magnitude (β = 0.7394, p = 0.07), while the coefficient on the individual norm rating is insignificant (β = 0.3597, p = 0.26). The AIC for the joint model is worse than for the model using only the average norm rating but it is better than the model that uses the individual norm ratings, while the BIC is worse for both. A Likelihood ratio test marginally rejects the model using the individual norms in favor of the combined model (p = 0.065), while not rejecting the model using the average norms (p = 0.25).

[25]Recall that these patterns were: (1) B requests a renegotiation, A accepts the request and then B rolls; (2) B does not make the request and then rolls anyway; (3) B does not make the request and does not roll.

The first column reports the average norm strength (from the norm elicitation task) for the sub-sample of B subjects who requested a renegotiation in the risky trust game. Similarly, the second column reports the average norm strength for B subjects who did not request a renegotiation and then did not roll in the risky trust game, and the third column reports the average norm for B subjects who did not request a renegotiation and then did roll in the risky trust game.

Each row in Table 6 uses the responses from the norm elicitation task to report the norm strength for "roll" when a request was accepted (ΔN^{Acc}), was rejected (ΔN^{Rej}) and the norm strength in the baseline where no negotiation was possible (ΔN^{Base}). In our theoretical analysis we assume that the norm will be the same in the case of choosing not to ask for renegotiation as it is when renegotiation is not possible. The numbers in each cell have only relative meaning, such that if a number is larger, then the norm is said to be stronger.

Table 6 shows that for subjects who requested a renegotiation in the risky trust game (column 1), the norm to roll is strongest when the request is accepted ($\Delta N^{Acc} = 1.15$) and weakest when it is rejected ($\Delta N^{Rej} = 0.46$). This is broadly consistent with what one might expect for subjects for whom the renegotiation outcome will matter for their roll decision – however, we will be more precise about this in Table 7 when we disaggregate this group of subjects. We focus here on Columns 2 and 3, which report the data we need to test the predicted norm pattern for those subjects who, in the risky trust game, either *don't ask and don't roll* or *don't ask and roll.*

For subjects with the behavioral pattern of *don't ask and don't roll* we stated that such a pattern would be consistent with the norms model if B has a weak norm in the baseline (small ΔN^{Base}), but a strong norm for rolling if a request were accepted (i.e. ΔN^{Acc} is large), and that the increased utility from switching behavior after a potential acceptance isn't that attractive (i.e. $N^{Acc}(R|X) - N^{Base}(D|X)$).

Table 6 shows that all of these assumptions are supported. First, 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), and it is comparable to the ΔN^{Acc} for the other two groups (Kruskal-Wallis and pairwise rank-sum tests, diff. n.s.). Second, ΔN^{Base} is directionally smaller than both other groups, with a significant difference compared to column 3 (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 that "do not ask" and "roll" (1.33). We also find that ΔN^{Rej} is significantly smaller than the corresponding norms for subjects in column 3, which is a natural consequence of ΔN^{Base} being small (and Assumptions 5 and 6).

The theoretical support for the pattern of *don't ask and roll* is the B player has a sufficiently strong norm that he will roll even if he asks and is rejected (and hence in equilibrium he would be rejected), i.e. that ΔN^{Rej} is large compared with the ΔN^{Rej} of other subjects. Note that as a consequence of Assumptions 5 and 6 we should also expect ΔN^{Base} and ΔN^{Acc} to be large.

In Table 6 we see that this is supported – this group of subjects has a large ΔN in all three scenarios. The difference is large and significant in ΔN^{Rej}, the key determinant for this pattern (ΔN^{Rej} = 1.27 vs 0.46 and 0.57; p = 0.01 compared to column 1, p = 0.09 compared to column 2, p = 0.02 compared to column 1 and 2 pooled). ΔN^{Base} is also substantially and significantly larger (ΔN^{Base} = 1.53 vs 0.96 and 0.78; p = 0.05 compared to column 1, p = 0.02 compared to column 2, p = 0.03 compared to column 1 and 2 pooled). ΔN^{Acc} is directionally larger, but not significantly so (ΔN^{Acc} = 1.47 vs. 1.15 and 1.20). Similarly, if we measure overall norm strength across all three scenarios by taking the sum of the ΔN values (i.e. we sum across the three cases of accept, reject and baseline and denote this as $\Sigma\Delta N$), we find that this group has a much stronger total norm strength ($\Sigma\Delta N$ = 4.27 vs 2.58 and 2.55; p = 0.04 compared to column 1, p = 0.05 compared to column 2, p = 0.03 compared to column 1 and 2 pooled). Hence the norm data for this group is consistent with subjects having such a strong norm that they don't want to ask for renegotiation because they worry about being rejected and receiving less norm utility when they roll anyway.

Table 7 about here.

Table 7 breaks down column (1) of Table 6, i.e. the subjects who requested renegotiation, based on the outcome of the renegotiation (accept or reject) and B's subsequent "roll" decision in the risky trust game. Columns (1) and (2) report the elicited norms for subjects that had their renegotiation request accepted; column (1) has subjects who ultimately rolled, while (2) has subjects that did not roll. Similarly, columns (3) and (4) report the elicited norms for subjects that had their renegotiation request rejected; column (3) has subjects who rolled, while (4) has subjects who did not roll. Note that since these patterns depend on A's action, we need to choose which

groups to focus on carefully so that we can distinguish between our two theoretical patterns for subjects that request renegotiation.

Column (1) provides clean evidence about the norms for subjects described by our *successful renegotiation* pattern. Recall that the on-equilibrium path of play is for B to request renegotiation, A to accept and B to rolls. Additionally, note that no other equilibrium gives this is pattern of play by B even if A takes a surprise action. The theoretical prediction for this behavior patterns is that ΔN^{Base} and ΔN^{Acc} are large and ΔN^{Rej} is small.

We find that ΔN^{Base} is fairly large, and not significantly different from the other columns (joint and pairwise comparisons n.s.). However, ΔN^{Rej} is much smaller than all other columns, with the differences significant (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). Further, when we take the difference in norm strength ($\Delta N^{Base} - \Delta N^{Rej}$) we find that this is much larger for column 1 than for the other subjects who renegotiated (0.88 versus 0.53 for columns 2 to 4 pooled, $p < 0.01$); we see the same difference comparing column 1 to all other subjects (0.88 versus 0.27 for all other subjects, $p < 0.01$).[26] Finally, we also see that ΔN^{Acc} is also fairly large – comparable to columns (3) and (4) and significantly larger than column (2) ($p = 0.05$). We take this together as supportive evidence that successful renegotiation is associated with particular characteristics about in the social norms for those subjects: ΔN^{Base} and ΔN^{Acc} large and ΔN^{Rej} small.

The second behavior pattern in the risky trust game, *ask expecting rejection,* we have to be a bit careful in selecting which group to focus on. One might naturally want to look at column (4) – subjects that were rejected and chose not to roll. While this set of behaviors is the on-equilibrium path of play for *ask expecting rejection*, it is also what B's are predicted to do in the other equilibrium if B is surprised by a rejection from A. Instead, to find evidence for *ask expecting rejection,* we focus on column (2), which is the off-equilibrium path for *ask expecting rejection*, after a surprise acceptance by A. In this path B asks for renegotiation (expecting a rejection, and planning not to roll), receives an acceptance by A, and then does not roll anyways because of a

[26]That is, when we compare to all other subjects we add in those who did not renegotiate in columns 2-4.

weak norm across the board. The theoretical prediction is therefore that all the ΔN's are small, especially ΔN^{Acc}.

As expected we find that ΔN^{Acc} is quite small (0.80) for this group. Additionally, ΔN^{Acc} is directionally smaller than all three other groups, and is significantly smaller than column (1), and marginally significantly smaller than the other three columns pooled ($p = 0.05$ and $p = 0.09$ respectively). We also find that ΔN^{Base} is smallest for this group, with significant difference compared column (3), column (4) and the other three columns pooled ($p < 0.05$ for each comparison). Additionally, ΔN^{Rej} is equally small compared to column (1), and is significantly smaller than column (3), column (4) and the other three columns pooled ($p = 0.01$, $p = 0.01$ and $p = 0.06$, respectively). Finally, if we use the sum of the ΔNs (ΣΔN), the total norm strength across the three renegotiation outcomes is significantly lower for Column (2) compared to 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 supportive evidence that subjects with behavioral patterns consistent with *ask expecting renegotiation* (given A's acceptance) also have the corresponding norm.

Although the pattern for column 3 isn't covered in one of our equilibrium patterns (since always rolling even if rejected should imply rejection, which should imply not asking) we do note that the observed norms are intuitive. If a subject with strong norms (and hence someone who intends to roll for any outcome) anticipates a sufficiently large change of acceptance, it could be a best response to request the renegotiation. This isn't unreasonable, given that we empirically observe A's accepting approximately half of requests. Hence, we should expect this group to have a strong norm across all three renegotiation outcomes – which is what we find. In particular, ΔN^{Rej} is largest for this group, with significant differences compared to Columns (1) and (2) ($p < 0.01$ for both comparisons).

In summary, our analysis of the norm strength by sub-groups of B players shows norm strength patterns that were predicted by the norms model. We now provide another analysis showing the connection between norm strength and behavior by using OLS regressions to predict the decision

to ask for renegotiation based on patterns in the social norm.[27] We create dummy variables for the different patterns of norm strength and use that to predict the behavior.

Specifically, since the major criteria for the norm revolve around various configurations of ΔN being large or small, 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: None above the median, Base only above, Base and Rej above, Acc only above, Acc and Base above, and all three above the median. We then use these norm patterns to predict the decision to ask for renegotiation in the risky trust game. The results are reported in Table 8. 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 made a promise.

Table 8 about here.

We find that norm strength significantly predicts the decision to request renegotiation. Subjects that have only ΔN^{Acc} as a stronger than the median norm (High A) are significantly more likely to ask for renegotiation (for both the full sample, and the subsample that promised). We also find marginal significance if we jointly test High A and High AB, i.e. subjects that just have ΔN^{Acc} or have both ΔN^{Acc} and ΔN^{Base} as stronger than the median norm ($p = 0.06$). This is consistent with the *successful renegotiation* equilibrium – subjects have norms that would justify A accepting, and are such that they would benefit from renegotiation. By contrast, subjects that have all three norms stronger than the median (High ABR) are 40% to 60% less likely to renegotiate, 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). Again, this is consistent with our theoretical analysis – having a strong norm after rejection makes renegotiation risky, since B may end up rolling anyways but with much lower utility.

We conclude our analysis by taking a step back and testing an unstated assumption of our norms model: that individuals care to comply with (any) norms and want to make choices that make social norms more important. In our game this comes from making a promise – which we

[27]We use a linear probability model since we are primarily using dummy explanatory variables, and with probit or logit regressions certain configurations of the explanatory variables perfectly predict the outcome.

have seen intensifies all of the social norms. Recall that in our game promising has no strategic value – all it can do is magnify the importance of the norm-compliance utility. Our final analysis tests whether those B players who make a promise are different from those who do not. Table 9 reports the average norm rating for those who made a promise and those who did not in the risky trust game.

Table 9 about here.

We see that B players who make a promise are those who think that making a promise will have the biggest effect on making the norms stronger – i.e. they 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 making norms stronger can be beneficial. This contrasts with many 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 disutilty from a breach in the agreement. Absent strategic motivations subjects have little reason to magnify the psychic costs of such a break.

Finally we also see that beliefs about the norms matter for A subjects' decision to play the game. Those A's that go "in" think that the norm to roll is stronger in most cases (a majority of the pairwise comparisons reported in the appendix are significant). This is quite 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.

VII. Discussion

The desire to uphold promise-keeping norms greases the wheels of interaction by creating trust. Norms establish a set of mutual expectations which parties rely on to interact in the presence of uncertainty and renegotiation. In this paper we do not treat the norms as unconditional with respect to the unforeseen contingencies and behaviors of the interacting parties. Rather, we assume that they change as uncertainty is resolved and by the very presence of an opportunity for, and outcomes achieved in, renegotiation. We show that they are able to explain much more of the behaviors that give rise to fulfilling an obligation than previously thought.

To do so, we present a model of social norm compliance in a risky trust game. We establish a set of assumptions about the norm that characterize how promises affect the norm to fulfill an agreement, how the norm is changed once unforeseen contingencies are resolved and is changed if a renegotiation request is accepted or rejected. We demonstrate that a utility is able to account for the pronounced effect of promises on behavior, explain why a second mover would make a promise in the first place and why a first mover would be willing to enter such a transaction.

We also show that renegotiation requests are not as frequent as a standard selfish model would predict and that a second mover's willingness to follow through on a promise is affected by whether the renegotiation is accepted or rejected. We demonstrate that such a pattern of behavior is again consistent with a utility model in which the social norm for promise keeping changes if requests for renegotiation are rejected. In particular we show that when second movers derive utility from complying with social norms to honor promises, we can explain why second movers make promises and why those promises change behavior. Further, a norms-based utility model is able to explain why second movers may *not* request a renegotiation and still choose to honor their promise nonetheless. Finally, we show that when renegotiation is rejected, then it is accompanied by a weakening of the promise-keeping norm which can explain why promises are not honored.

There are at least two other prominent behavioral models of contracts (Hart and Moore 2008 and Herweg and Schmidt 2014) that have been used to explain renegotiation behavior in other experimental settings (Fehr et al. 2011 and Bartling and Schmidt 2015). In the Hart and Moore model the initial agreement sets a reference point for what fair treatment consists of under the agreement. In the Herweg and Schmidt (2014) model renegotiated outcomes may be interpreted as losses when they are worse for one party than the initial agreement would indicate. How well can these other behavioral models of renegotiation apply to our setting? We show that two important behaviors in the renegotiation treatment cannot be explained by the reference-point based models: that B's will be more likely to roll after an acceptance than a rejection, and that many B's will choose not to renegotiate. Similarly, outcome-based fairness models cannot explain the impact of promises, B's willingness to roll, or B's choice not to ask for renegotiation. We have shown how all of these outcomes can be explained within a social norms framework given appropriate assumptions about the shape of the norm, but that other behavioral models typically applied to renegotiation cannot.

One implication of our study comes from the subprime mortgage crisis. At its peak in 2012 12.8 million U.S. homeowners with a mortgage (28.6%) owed at least 25% more on their home than the property was worthy (so called "seriously underwater" borrowers). Though some strategically defaulted, the vast majority continued to pay off their mortgage. Anecdotal evidence from interviews with homeowners of underwater mortgages suggests that part of a homeowner's decision to repay their mortgage was motivated by a norm to honor the promise of repayment – they expressed unease when they failed to honor their repayment promise.[28] Further, and perhaps surprisingly, they also expressed unease with requesting to renegotiate terms with the bank; they cited making a promise to repay the loan as the reason. This suggests that making promises may have the surprising and unintended effect of inhibiting renegotiation utilization.

Further, banks seem to be aware of these promise-keeping norms[29] and, in turn, rely on these norms as a reason to reject requests for renegotiation. If interacting parties believe that there is a strong norm of promise-keeping, then they may be less likely to agree to a renegotiation and, this in turn, could further diminish the incentive to request the renegotiation in the first place. Awareness of and reliance on these promise keeping norms may shed light on the low response rate by banks to the 2009 Homeowner's Affordability plan[30]; a plan that provided incentives to banks to renegotiate underwater mortgages (Agarwal et al. 2012). Testifying before the Senate Committee on Banking, Housing and Urban Affairs in 2009, Economist Willens stated that one reason banks didn't renegotiate terms was "the possibility that the borrower would have repaid the loan without any assistance from the lender…An investor would view assistance given …as 'wasted' money." Such statements may stem from a (not unfounded) belief in the normative power of the promissory note to dissuade borrowers' from strategic default.[31]

[28]A widely cited study of defaults, conducted by Experian and the consulting firm Oliver Wyman's, estimated that only 17% of 2008 defaults were strategic.

[29]Former Treasury Secretary Paulson Jr. declared "any homeowner who can afford his mortgage payment but chooses to walk away from an underwater property is simply a speculator — and one who is not honoring his obligation." (For the full article, see http://www.nytimes.com/2010/01/10/magazine/10FOB-wwln-t.html)

[30] On 18 February 2009, U.S. President Barack Obama announced a $73 billion program to help up to nine million homeowners avoid foreclosure, which was supplemented by $200 billion in additional funding for Fannie Mae and Freddie Mac to purchase and more easily refinance mortgages.

[31]For homeowners who did request new terms on their mortgages and banks who agreed to negotiate new terms, Agarwal et al. (2012) find that the "redefault" rate on renegotiated mortgages was significantly lower than that of the comparison group. This is certainly due to the improved terms of the loan, but there may also be some increased willingness to avoid default *because* the bank is willing to renegotiate the loan.

A second implication of our study has to do with the informal transmission of norms. If people have a strong enough preference to comply with the norm, then they may choose to avoid situations where the norms would negatively impact their utility (either because they would honor the norm even when they know that others would not or because they know that they are honoring the norm for someone who has not been kind to them). This implies that signaling of the norm is noisy or does not happen at all. The result is that transmission of the norm can be weakened.

VIII. Conclusion

In this paper we focus on whether and how agreements are fulfilled and argue that a social norms frame-work can best explain the suite of behavior patterns that give rise to fulfilling an agreement. Our main contribution is to be able to explain a host of behaviors with one model. We are able explain why promises are made; we can explain the willingness to fulfill promises even when it is costly to do so; we are able to explain how renegotiation success or failure affects the propensity of individuals to fulfill promises, and we can explain why someone would choose to request or not request a renegotiation of an initial commitment after uncertainty is resolved. The norms framework is both flexible and powerful – making clear predictions based on easily observable features of the social norm. Our results add to our understanding of how norms affect renegotiation, offer important refinements to our knowledge of how to structure contracts and provide a novel extension of the social norms and promises literature by exploring how norms for promise keeping change as a function of uncertainty and in response to actions taken by interacting parties.

References

Andreoni, J., & Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5), 1607-1636.

Bartling, B., & Schmidt, K. M. (2015). Reference points, social norms, and fairness in contract renegotiations. *Journal of the European Economic Association*, 13(1), 98-129.

Burks, S. V., & Krupka, E. L. (2012). A multimethod approach to identifying norms and normative expectations within a corporate hierarchy: Evidence from the financial services industry. *Management Science*, 58(1), 203-217.

Carlton, D. W. (1986). The Rigidity of Prices.*The American Economic Review*, 637-658.

Camerer, C. F., Fehr, E. (2004). Measuring social norms and preferences using experimental games: A guide for social scientists. in Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, eds. Foundations of Human Sociality-Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies (Oxford University Press, Oxford, UK), 55–95

Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74(6), 1579-1601.

Charness, G., & Grosskopf, B. (2004). What makes cheap talk effective? Experimental evidence. *Economics Letters*, 83(3), 383-389.

D'Adda, G., Drouvelis, M., & Nosenzo, D. (2015). Norm elicitation in within-subject designs: Testing for order effects. (No. 2015-02). *CeDEx Discussion Paper Series*.

Dufwenberg, M., Gächter, S., & Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior*, 73(2), 459-478.

Erkut, H., Nosenzo, D., & Sefton, M. (2015). Identifying social norms using coordination games: Spectators vs. stakeholders. *Economics Letters*, 130, 28-31.

Eggleston, K., Posner, E. A., & Zeckhauser, R. (2000). The design and interpretation of contracts: why complexity matters. *Northwestern. University Law Review,95*, 91.

Ellingsen, T., & Johannesson, M. (2004). Promises, threats and fairness. T*he Economic Journal*,114(495), 397-420.

Falk, A., & Kosfeld, M. (2006). The hidden costs of control. *American Economic Review*, 96(5): 1611-1630.

Federal Reserve Bank of Chicago; et al. (2012). Policy Intervention in Debt Renegotiation: Evidence from the Home Affordable Modification Program, 2012. http://journalistsresource.org/studies/economics/real-estate/policy-intervention-debt-renegotiation-evidence-home-affordable-modification-program#sthash.GCBEj1FR.dpuf

Fehr, E., & Falk, A. (1999). Wage rigidity in a competitive incomplete contract market. *Journal of Political Economy*, 107(1), 106-134.

Fehr, E., Hart, O. D., & Zehnder, C. (2011). How do informal agreements and renegotiation shape contractual reference points? (No. w17545). *National Bureau of Economic Research*.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3), 817-868.

Feltovich, N., & Swierzbinski, J. (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, S., Nosenzo, D., & Sefton, M. (2013). Peer effects in pro-social behavior: Social norms or social preferences? *Journal of the European Economic Association*, 11(3), 548-573.

Hart, O., & Moore, J. (2008). Contracts as reference points. *The Quarterly Journal of Economics*, 123(1), 1-48.

Hart, O., & Moore, J. (2008). Contracts as reference points. *The Quarterly Journal of Economics*, *123*(1), 1-48.

Hoffman, S. D., & Duncan, G. J. (1988). Multinomial and conditional logit discrete-choice models in demography. *Demography*, 25(3), 415-427.

Hoppe, E. I., & Schmitz, P. W. (2011). Can contracts solve the hold-up problem? Experimental evidence. *Games and Economic Behavior*, 73(1), 186-199.

Iyer, R., & Schoar, A. (2015). Ex post (in) efficient negotiation and breakdown of trade. *The American Economic Review*, 105(5), 291-294.

Kessler, J. B., & Leider, S. (2012). Norms and contracting. *Management Science*, 58(1), 62-77.

Kimbrough, E. O., & Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3), 608-638.

Kreps, D. (1990). M. 1990. Corporate culture and economic theory. *Perspectives on Positive Political Economy*,90, 109-10.

Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3), 495-524.

Krupka, E. L., Leider, S., & Jiang, M. (2016). A meeting of the minds: informal agreements and social norms. *Management Science*, 63(6): 1708-1729.

Lambert, N., & Shoham, Y. (2009, July). Eliciting truthful answers to multiple-choice questions. *Proceedings of the 10th ACM conference on Electronic commerce*, (pp. 109-118). ACM.

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

Macaulay, S. (1963). Non-contractual relations in business: A preliminary study. *American Sociological Review*, 55-67.

McFadden, D. (1974). Analysis of Qualitative Choice Behavior. Zarembka, P.(ed.): Frontiers in Econometrics.

Miettinen, T. (2013). Promises and conventions–An approach to pre-play agreements. *Games and Economic Behavior*, 80, 68-84.

Rigdon, M. (2009). Trust and reciprocity in incentive contracting. *Journal of Economic Behavior & Organization*, 70(1), 93-105.

Scott, R. E. (2003). A theory of self-enforcing indefinite agreements. *Columbia Law Review*, 103, 1641.

Sliwka, D. (2007). Trust as a signal of a social norm and the hidden costs of incentive schemes. *The American Economic Review*, 97(3), 999-1012.

Tirole, J. (1999). Incomplete contracts: Where do we stand? *Econometrica*, 67(4), 741-781.

Vanberg, C. (2008). Why do people keep their promises? an experimental test of two explanations. *Econometrica*,76(6), 1467-1480.

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

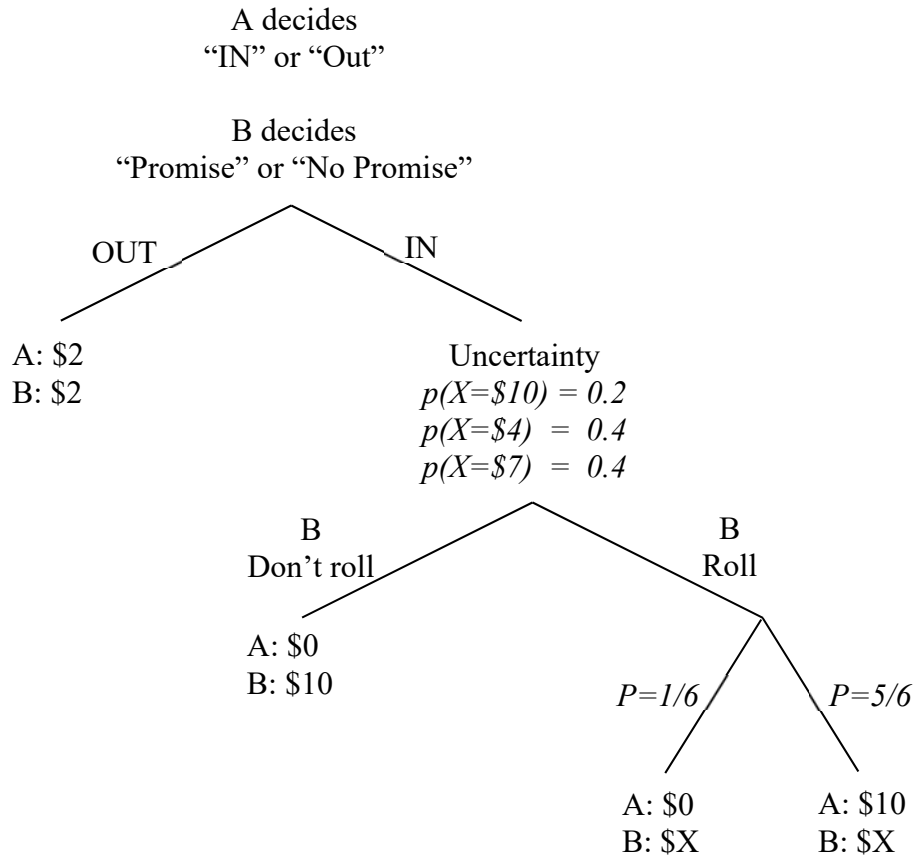

Figure 1: Risky trust game

<table>
<tr><td colspan="7">Risky Trust Game</td><td>Norm Elicitation Task</td></tr>
<tr><td>Timeline</td><td>T = 0</td><td>T = 1</td><td>T = 2</td><td>T = 3</td><td>T = 4</td><td>T = 5</td><td>T = 6</td></tr>
<tr><td>Stages</td><td>Stage 1</td><td>Stage 2</td><td colspan="2">Stage 3</td><td>Stage 4</td><td>Stage 5</td><td>Stage 6</td></tr>
<tr><td>Subject Role</td><td>Choice stage</td><td>Information stage (all see)</td><td colspan="2">Renegotiation (treatment only)</td><td>B ROLL choice</td><td>Information stage (all see)</td><td>Ratings task</td></tr>
<tr><td>A</td><td>In / Out</td><td rowspan="2">A’s IN/OUT revealed

B’s payoff for rolling is revealed to be:
X= $10, $7, or $4</td><td></td><td>Accept/Decline</td><td></td><td rowspan="2">All see results</td><td rowspan="2">Rating norms for 3 different situations</td></tr>
<tr><td>B</td><td>Promise / No Promise</td><td>Ask/Don’t Ask</td><td></td><td>Roll/Don’t Roll</td></tr>
</table>

Note: p(X=$4)=0.4; p(X=$7)=0.4; p(X=$10)=0.2

Figure 2: Overview of the experimental design.

B makes a promise, X =$4 and B decides to NOT ROLL the die	○ Very Socially Appropriate ○ Socially Appropriate ○ Somewhat Socially Appropriate ○ Somewhat Socially Inappropriate ○ Socially Inappropriate ○ Very Socially Inappropriate	B makes a promise, X=$4 and B decides to ROLL the die	○ Very Socially Appropriate ○ Socially Appropriate ○ Somewhat Socially Appropriate ○ Somewhat Socially Inappropriate ○ Socially Inappropriate ○ Very Socially Inappropriate
B makes a promise, X=$7 and B decides to NOT ROLL the die	○ Very Socially Appropriate ○ Socially Appropriate ○ Somewhat Socially Appropriate ○ Somewhat Socially Inappropriate ○ Socially Inappropriate ○ Very Socially Inappropriate	B makes a promise, X =$7 and B decides to ROLL the die	○ Very Socially Appropriate ○ Socially Appropriate ○ Somewhat Socially Appropriate ○ Somewhat Socially Inappropriate ○ Socially Inappropriate ○ Very Socially Inappropriate

Figure 3: Excerpt of the norm ratings table from the *baseline vignette* in the norm elicitation task.

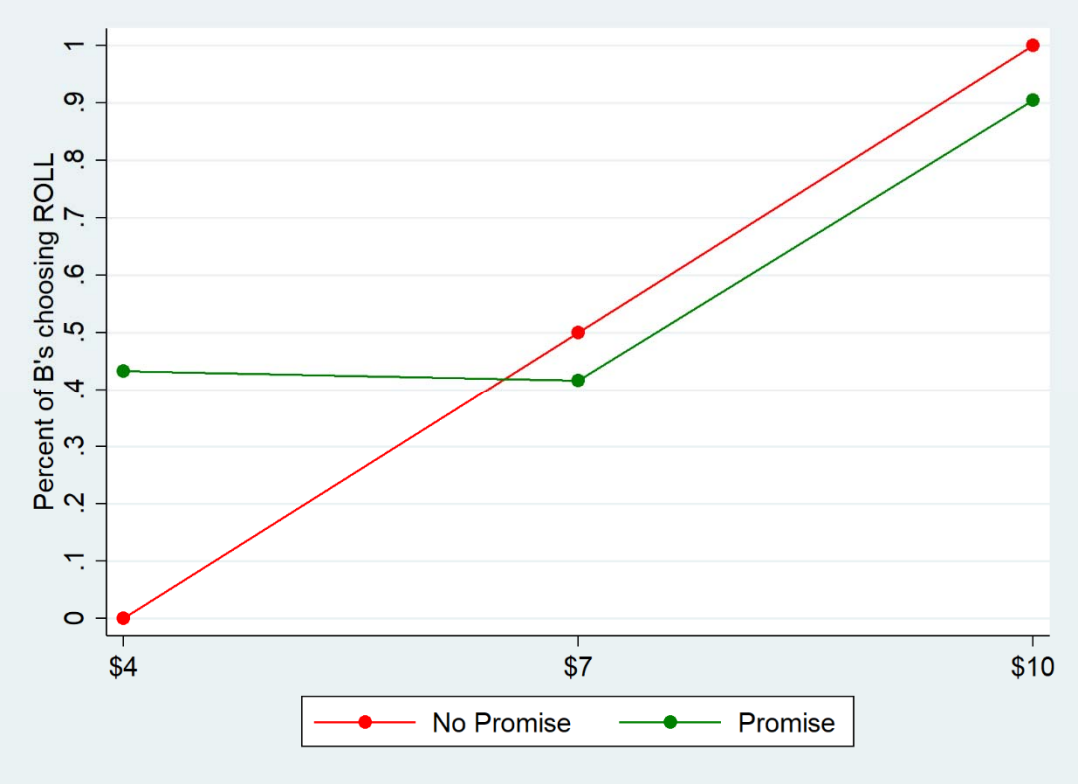

Figure 4: B's decision to "roll" by "promise" for different dollar values of X ($4, $7, and $10) for pooled data from the baseline and renegotiation treatment of the risky trust game.

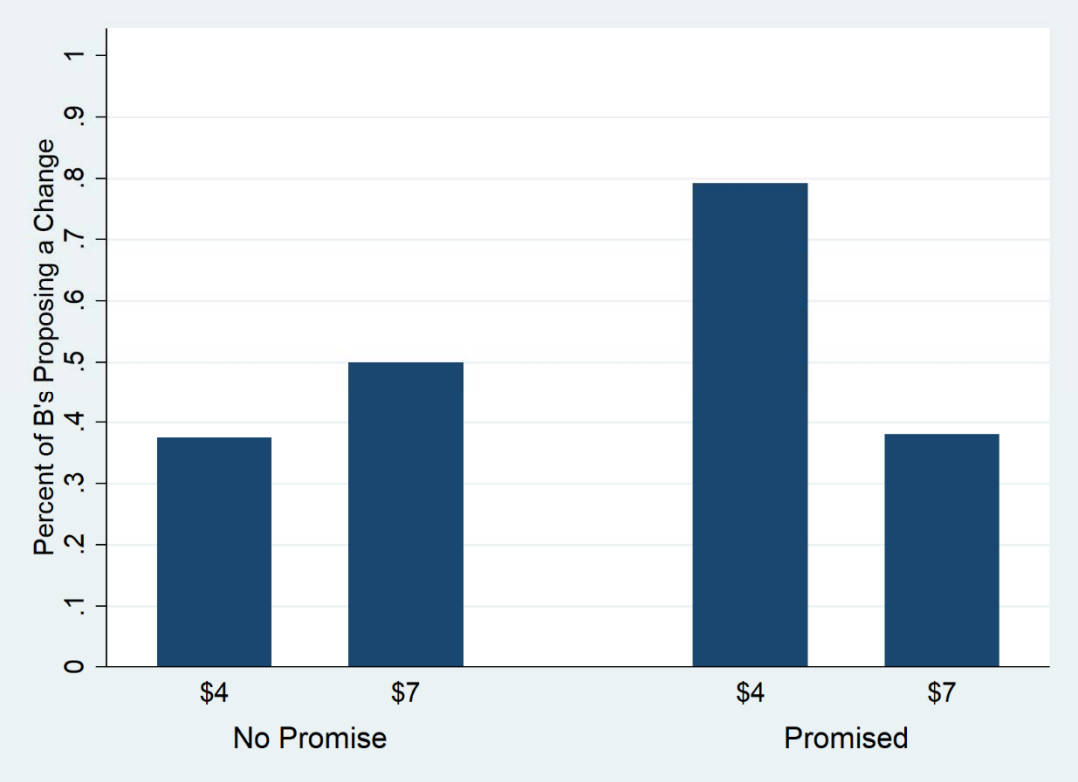

Figure 5: Percent of B's who request renegotiation by "promise" and by X = $4 or X = $7 in the renegotiation treatment of the risky trust game.

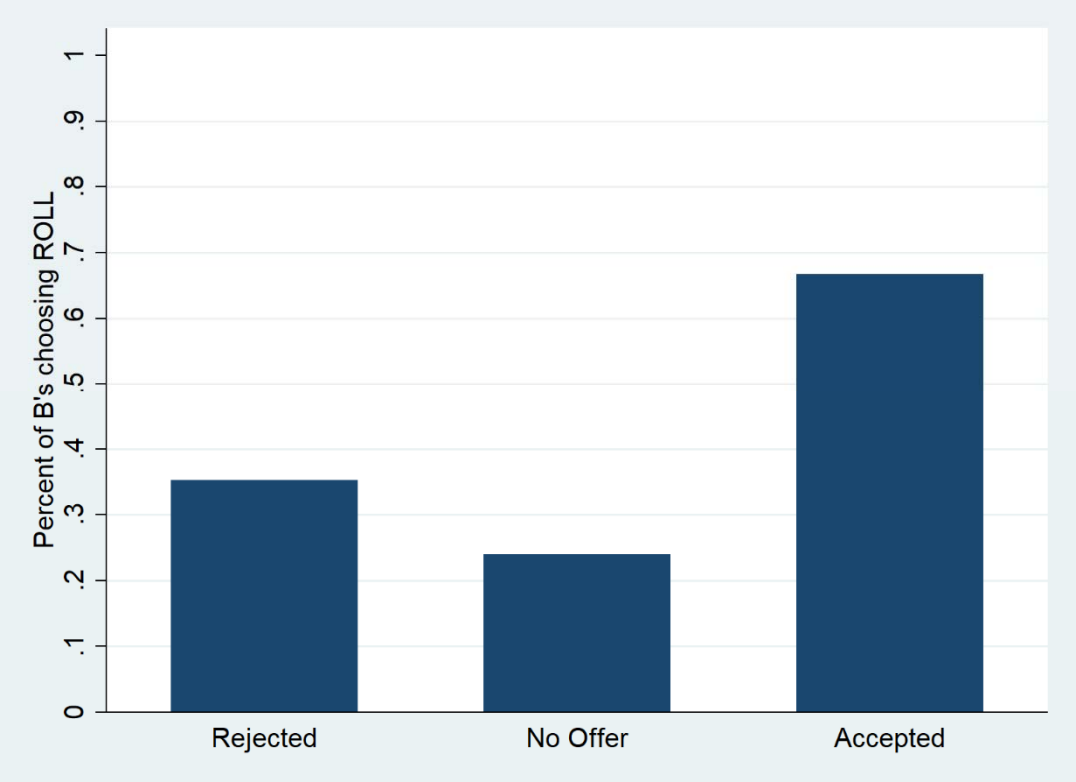

Figure 6: Percent of B's choosing to roll by the outcome of the renegotiation request in the renegotiation treatment of the risky trust game.

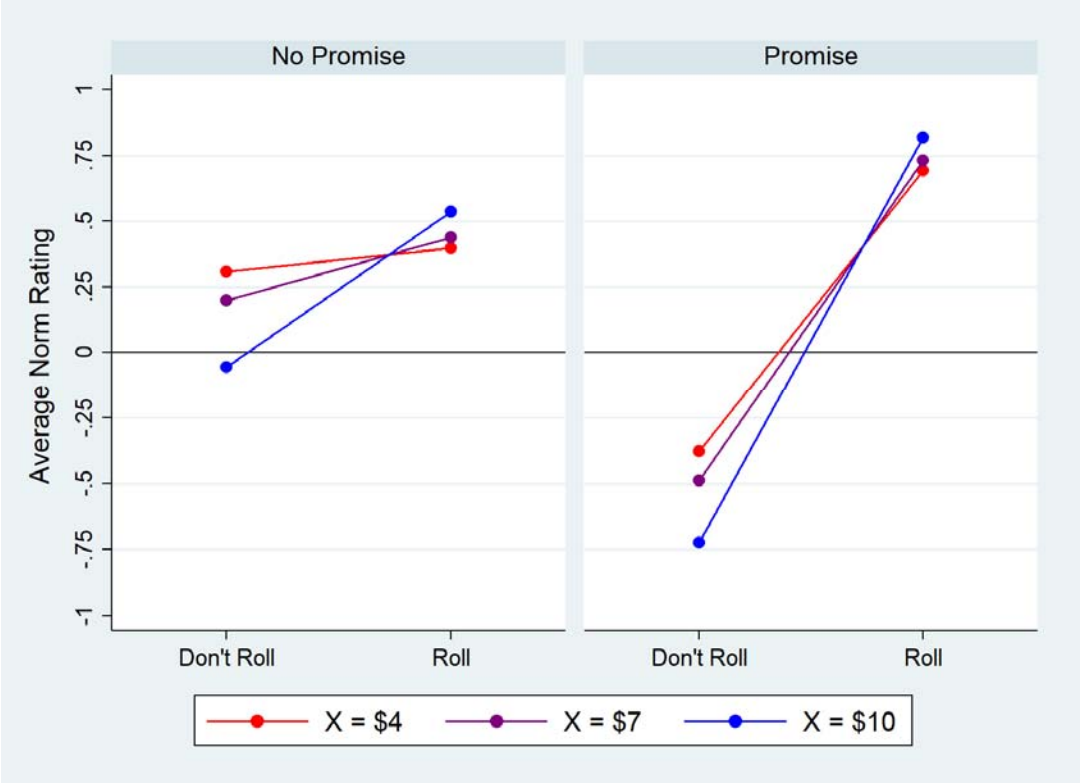

Figure 7: Mean norm ratings for "roll" and "don't roll" in the *baseline vignette* (where renegotiation is not an option) from the norm elicitation task.

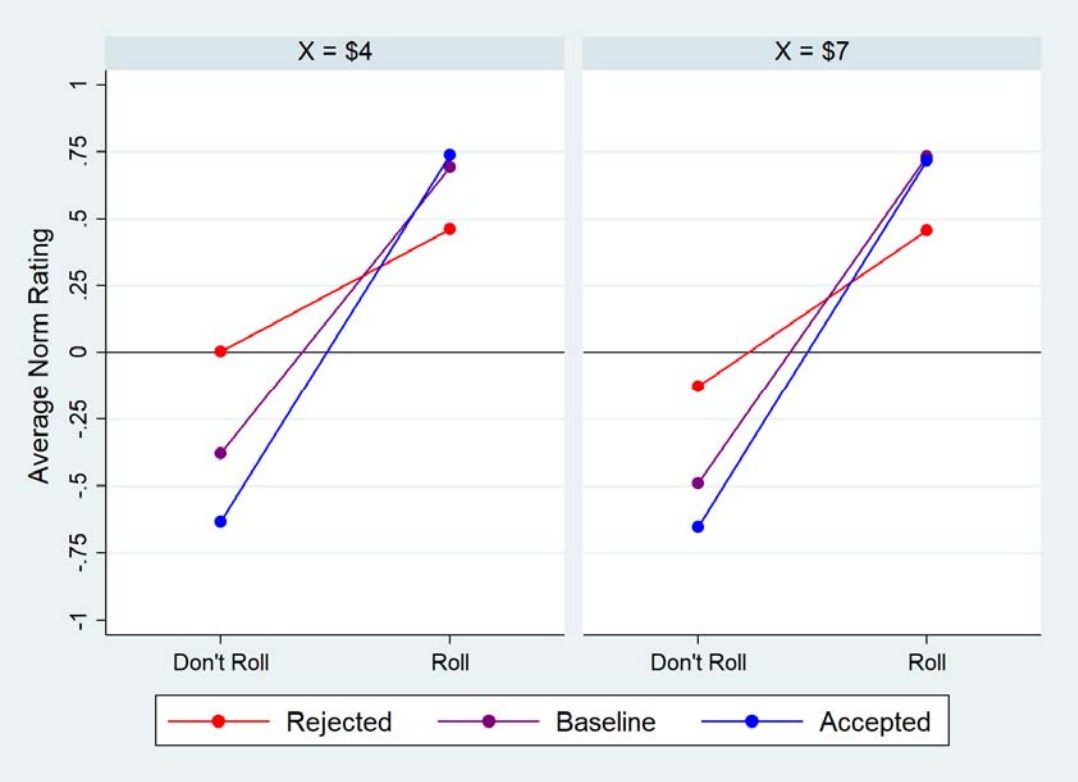

Figure 8: Mean norm ratings for "roll" and "don't roll" when B has made a promise. Ratings reported for two renegotiation outcomes (rejected and accepted) and for the baseline (no negotiation possible) from the norm elicitation task.

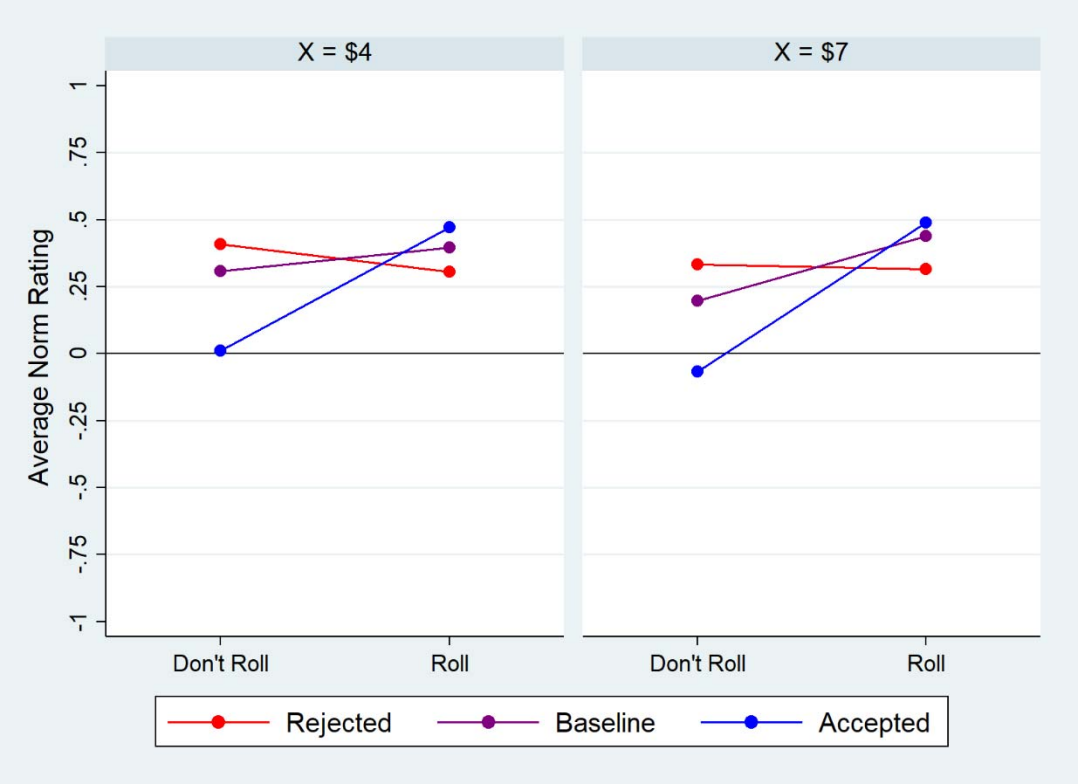

Figure 9: Mean norm ratings for "roll" and "don't roll" when B has not made a promise. Ratings reported for two renegotiation outcomes (rejected and accepted) and for the baseline (no negotiation possible) from the norm elicitation task.

	A Receives	B Receives
A Chooses "OUT"	$2	$2
A Chooses "IN", B Chooses "NOT ROLL"	$0	$10
A Chooses "IN", B Chooses "ROLL", ROLL=1	$0	X = $
A Chooses "IN", B Chooses "ROLL", ROLL= 2,3,4,5, or 6	$10	X = $

Note: Subjects were given all the information provided in the table. They also know that X could take the value $4, $7 or $10 and the probabilities associated with each outcome. Those were p(X=$4)=0.4; p(X=$7)=0.4; p(X=$10)=0.2.

Table 1: Payoff table for subjects in the baseline of the risky trust game.

	Proposal if X = $4		**Proposal if X = $7**	
Possible Outcomes if A Accepts Proposal	**A receives**	**B receives**	**A receives**	**B receives**
B chooses DON'T ROLL	$0	$10	$0	$10
B chooses ROLL, die =1	$0	$7	$0	$8.50
B chooses ROLL, die =2	$0	$7	$0	$8.50
B chooses ROLL, die=3	$0	$7	$10	$8.50
B chooses ROLL, die=4	$10	$7	$10	$8.50
B chooses ROLL, die=5	$10	$7	$10	$8.50
B chooses ROLL, die=6	$10	$7	$10	$8.50
Possible Outcomes if A Rejects Proposal	**A receives**	**B receives**	**A receives**	**B receives**
B chooses DON'T ROLL	$0	$10	$0	$10
B chooses ROLL, die =1	$0	$4	$0	$7
B chooses ROLL, die =2	$10	$4	$10	$7
B chooses ROLL, die=3	$10	$4	$10	$7
B chooses ROLL, die=4	$10	$4	$10	$7
B chooses ROLL, die=5	$10	$4	$10	$7
B chooses ROLL, die=6	$10	$4	$10	$7

Table 2: Payoff table if the renegotiation is accepted or rejected in the renegotiation treatment of the risky trust game.

Observed Behavioral Pattern			% of Subjects		Consistent with Theory	
B Reneg	A Response	B Roll Decision	All	Promised	Primary Path of Play	If A Deviates
Yes	Accept	Roll	17.5%	22.2%	Successful Renegotiation	-
Yes	Accept	Don't Roll	8.8%	6.7%	-	Ask expecting Rejection
Yes	Reject	Roll	10.5%	11.1%	-	-
Yes	Reject	Don't Roll	19.3%	20.0%	Ask expecting Rejection	Successful Renegotiation
No	-	Roll	10.5%	13.3%	Don't Ask & Roll	-
No	-	Don't Roll	33.3%	26.7%	Don't Ask & DR	-

Table 3: Observed frequencies of each pattern of play when renegotiation is possible.

	Norm ratings for different values of X					
	X = $7			X = $4		
Promise	1.37	1.22	0.58	1.37	1.07	0.46
No Promise	0.55	0.24	-0.02	0.46	0.09	-0.10

Table 4: Testing the inequality relationship implied by Assumptions 5 and 6 using averaged responses from the norm elicitation task.

Variables	Using the Average Norm	Using the Individual Norm
Monetary Payoff	0.3793***	0.3522***
	(0.0863)	(0.0829)
Norm of Action	1.0865***	0.8246***
	(0.2691)	(0.2163)
N	107	107
Log Likelihood	-59.52116	-60.56725
AIC	123.0423	125.1345
BIC	128.388	130.4801

Table 5: Conditional logistic regression where the dependent variable is the "roll" decision in the risky trust game and norms are proxied with average or individual responses from the norm elicitation task.

Norm Strength for different norm scenarios	**Behavior of B Player in Risky Trust Game**		
	Requested Renegotiation	**No Request & Don't Roll**	**No Request & Roll**
	(1)	(2)	(3)
ΔN^{Acc}	1.15	1.20	1.47
ΔN^{Base}	0.96	0.78	1.53
ΔN^{Rej}	0.46	0.57	1.27

Table 6: Norm strength calculated from ratings in the norm elicitation task for subgroups of B players in the risky trust game: Those who requested, those who did not request and didn't roll and those who did not request and did roll.

Norm Strength for different norm scenarios	Behavior of B Players that Requested Renegotiation in Risky Trust Game			
	A Accepted and B Rolled	A Accepted And B Did Not Roll	A Rejected and B Rolled	A Rejected and B Did Not Roll
	(1)	(2)	(3)	(4)
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

Table 7: Norm strength calculated from ratings in the norm elicitation task for subgroups of B players who asked for renegotiation in the risky trust game conditioning on A's response (accepted or rejected).

VARIABLES	All B (1)	If Promise (2)	All B (3)	If Promise (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 between A and ABR	0.580***	0.529***		
	(0.177)	(0.189)		
Diff between 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

Dependent variable is decision to ask for renegotiation. Columns 1 and 3 include all subjects in renegotiation treatment where A chose in and X < 10. Columns 2 and 4 additionally restrict to subjects that promised. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Linear probability regression where the dependent variable is the renegotiation decision (accepted or reject) in the risky trust game.

Difference in Norm Strength	Behavior of B Player in Risky Trust Game	
	Promise	No Promise
	(1)	(2)
$\Delta N_4^P - \Delta N_4^{NP}$	1.02	0.72
$\Delta N_7^P - \Delta N_7^{NP}$	1.12	0.70
$\Delta N_{10}^P - \Delta N_{10}^{NP}$	1.12	0.80

Table 9: Difference in the norm strength between B players who made a promise and those who did not in the risky trust game.

Appendix A: Additional Theory Details

In this appendix we provide additional details showing that the various patterns of behavior that we observe in our experiment can be supported within our social norms framework.

Main Treatment (without renegotiation)

We first show that, given Assumptions 1-4, for any norm function N() there are values of η such that (some) B's will promise, B's that promise will Roll more than B's that don't, and (some) A's will choose In.

B's decision to Promise increases the likelihood of Rolling

To demonstrate that making a promise will increase the likelihood of choosing Roll, we show that there is a (non-empty) interval of values for η such that B will choose to Roll if he has promised, and will choose Don't Roll if he has not promised.

If B chooses to Roll, he receives $U_B(R) = X + \eta*N(R|X)$, while if he chooses Don't Roll he receives $U_B(D) = 10 + \eta*N(D|X)$. 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 $\eta*\Delta N_X \geq (10 - X)$. This is always true if X = 10, hence the more interesting cases are X = 4 or 7.

For X = 7, we need $\eta*\Delta N^P_7 \geq 3$ and $\eta*\Delta N^N_7 < 3$, or $\eta \in [3/\Delta N^P_7, 3/\Delta N^N_7)$. Similarly, for X = 4 the promise would change behavior if $\eta \in [6/\Delta N^P_4, 6/\Delta N^N_4)$. Given Assumptions 3 and 4, in both cases the intervals are non-empty.

B's are willing to Promise

Next, we want to ask whether an individual would be willing to make such a promise. If a player B is willing to roll even without a promise, then making a promise is straightforward. Similarly, if player B doesn't 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. First, suppose we fix X. Then a player B can choose to make a promise and Roll, receiving $U_B = X + \eta*N(R|P)$, or not promise and Don't Roll, receiving $U_B = 10 + \eta*N(D|N)$. He will then make the promise if $\eta*(N(R|P) - N(D|N)) > 10-X$. Since $\Delta N^P_X > N(R|P) - N(D|N) > \Delta N^N_X$ it is possible to find such a value of η for any X, given our maintained assumptions.

However, player B's have to choose to promise before they know what X is. Therefore, they will only be willing to promise if their 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 that are

consistent with B's promising, so we discuss here one example to demonstrate that a social norms model can justify promise making. 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 promise he would roll if X = 7 or 10 but not if X = 4. Then to be willing to promise he needs to have U(Promise) > U(No Promise), i.e.:

$$0.2*(\eta[N(R|P,10) - N(R|N,10)]) + 0.4*(\eta[N(R|P,7) - N(R|N,7)]) +$$

$$0.4*(\eta[N(R|P,4) - N(D|N,4)]-6) \geq 0$$

The first two terms are clearly positive, so a sufficient condition is for the last term to be positive. That is, we want $\eta[N(R|P,4) - N(D|N,4)] \geq 6$, or $\eta \geq 6/(N(R|P,4) - N(D|N,4))$. We also need to match our assumption of rolling for X = 4 if he promised, and not rolling if he didn't. By Assumption 4 our above assumption ensures that $\eta*\Delta N^{P}_{4} \geq 6$, and by Assumption 3 we have $6/\Delta N^{N}_{4} > 6/(N(R|P,4) - N(D|N,4))$. Therefore, there is an interval where if $\eta \in [6/(N(R|P,4) - N(D|N,4)), 6/\Delta NN4)$ then the player would be willing to make a promise. One can similarly construct examples for other patterns of rolling behavior.

A's are willing to choose In

If A chooses Out he will get \$2, while if he choose In he will get (on average) \$8.33 if player B rolls, and 0 if player B doesn't roll. Therefore, Player A will be willing to choose IN as long as the probability that B will roll is at least 24%. 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. Therefore, player 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.

Renegotiation Treatment

In this section we identify what conditions, in addition to Assumptions 1 to 6, would support various renegotiation outcomes. 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.

Successful Renegotiation

A successful renegotiation consists of: B asking, A accepting, and B rolling. Since accepting renegotiation lowers player A's payoff, he will only accept the renegotiation if it would make a difference for player B's decision to roll. Therefore, we need that player B would ROLL if A accepts, i.e. $\eta*\Delta N^{Acc}_{X} \geq ½(10-X)$, and that he would NOT ROLL if A rejects, i.e. $\eta*\Delta N^{Rej}_{X} < (10-X)$. Finally, we need B to prefer to ask, given that A will accept. B will receive the following utilities:

Asks: $U(\text{Ask, Roll}) = \frac{1}{2}(10+X) + \eta N^{Acc}(R|X)$

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

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

By Assumption 6 asking for renegotiation dominates simply rolling, so one sufficient condition for successful renegotiation is if $\eta * \Delta N^{Base}_X \geq (10-X)$ and $\eta * \Delta N^{Rej}_X < (10-X)$, or equivalently $\eta \in [(10+X)/\Delta N^{Base}_X, (10-X)/\Delta N^{Rej}_X)$, which is always non-empty. If B would choose Don't Roll in the base case, then the main restriction we need is that asking is preferable to not rolling, i.e. that $\eta \geq \frac{1}{2}(10-X)/[N^{Acc}(R|X) - N^{Base}(D|X)]$. By Assumptions 5 and 6 a sufficient condition is that $\eta \geq \frac{1}{2}(10+X)/\Delta N^{Base}_X$. Therefore this scenario can be supported if $\eta \in [\frac{1}{2}(10+X)/\Delta N^{Base}_X, (10-X)/\Delta N^{Rej}_X)$, which is always a non-empty interval given our assumptions. Note that this interval includes the previous one (associated with rolling in the base case). This interval is largest when ΔN^{Base} is large and ΔN^{Rej} is small.

Not asking for Renegotiation, and Rolling

Here we look at a case where B will Roll in the base case, and also prefers not to ask for renegotiation. In this case it has to be that A will reject renegotiation – since (given our assumptions) B would have to also Roll if accepted, and rolling after acceptance is always preferred to just rolling. A will only reject the renegotiation if it won't make a difference to B's action. Therefore, for this case we need that B will roll in all cases (accept, not ask, and reject), and thus $\eta * \Delta N^{Rej}_X \geq 10-X$. Comparing asking or not, B will receive the following utility:

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

By Assumption 5 not asking for renegotiation will always be preferred. Therefore, as long as $\eta \geq (10-X)/\Delta N^{Rej}_X$ the social norms model can explain B not asking for renegotiation and simply rolling. This requirement is weakest (and therefore allows the most values of η) when ΔN^{Rej} is large.

Not asking for Renegotiation, and Not Rolling

Here we look at a case where B will chose Don't Roll in the base case, and also prefers not to ask for renegotiation. In contrast to the previous case, here is has to be that A would accept renegotiation – if A were going to reject then B would always prefer to ask, since that would reduce the disutility from not rolling. For A to accept renegotiation it must be that B will roll if A accepts, and will not roll if A rejects. Therefore we have $\eta \geq \frac{1}{2}(10-X)/\Delta N^{Acc}_X$ and $\eta < (10-X)/\Delta N^{Base}_X$. Comparing asking or not, B will receive the following utility:

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

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

For B to prefer not to ask we need: $(10\text{-}X)/2 > \eta(N^{Acc}(R|X) - N^{Base}(D|X))$. This is possible given our earlier assumptions, since $(N^{Acc}(R|X) - N^{Base}(D|X)) = \eta(N^{Acc}(R|X) - N^{Base}(R|X) + \Delta N^{Base}{}_X)$ and $\Delta N^{Acc}{}_X - \Delta N^{Base}{}_X = (N^{Acc}(R|X) - N^{Base}(R|X)) + (N^{Base}(D|X) - N^{Acc}(D|X))$, with both terms positive. Therefore, we need $\eta \in [\frac{1}{2}\ (10\text{-}X)/\Delta N^{Acc}{}_X, \frac{1}{2}\ (10\text{-}X)/(N^{Acc}(R|X) - N^{Base}(D|X))\ )$. This interval is largest when ΔN^{Acc} is large and $(N^{Acc}(R|X) - N^{Base}(D|X)$ is small.

Ask for Renegotiation expecting Rejection

Here we look at a case where B's ask for renegotiation expecting (and in fact preferring) rejection. For A to reject renegotiation, it has to be that accepting won't change B's behavior – in this case that B will choose Don't Roll in both cases. Additionally, this means that B has to choose Don't Role in the base case as well. Therefore, in considering whether to ask he is comparing the following utilities:

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

Ask is clearly preferred given Assumption 6, since $N^{Rej}(D|X) > N^{Base}(D|X)$. Furthermore, the rest of Assumption 6 indicates that B prefers his renegotiation request be rejected:

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

Rejection is preferred since $N^{Rej}(D|X) > N^{Acc}(D|X)$.

Therefore, this case obtains if $\eta < \frac{1}{2}\ (10\text{-}X)/\Delta N^{Acc}{}_X$. This is easiest to satisfy when ΔN^{Acc} is small.