A Meeting of the Minds: Contracts and Social Norms

Erin L. Krupka (School of Information, University of Michigan & IZA)

Stephen Leider (Ross School of Business, University of Michigan)

Ming Jiang (School of Information, University of Michigan)

October 4, 2011[1]

<<NOT FOR CIRCULATION WITHOUT PRIOR APPROVAL OF THE CO-AUTHORS, PLEASE>>

Abstract

In this paper we demonstrate that incomplete contracts induce specific social norms of obligation. We employ a new experimental method to measure social norms directly via an incentive-compatible mechanism that exploits a key feature of norms: they are collectively help perceptions of the appropriateness of behavior. We find that handshake agreements substantially change the social norm in three ways. First, taking the promised action becomes substantially more appropriate, and all other actions become less appropriate. Second, the handshake agreement increases the consensus across individuals about which action is the most appropriate. Third, in the Bertrand Game the handshake agreement replaces a norm of risk minimization with a norm of obligation. Our results shed new light on one mechanism by which incomplete contracts persist and can outperform complete contracts. Finally we combine choice data for these games with the social norms elicited using the incentive compatible norm elicitation technique to predict changes in behavior across conditions and games. We show that a utility model that includes social norms as an additional motivation does much better predicting behavior than models which include only monetary utility and social preferences for fairness.

JEL Codes: C93, D23

---

[1] We wish to thank Sally Meyers and Felicia Kessler, Caitlin Holman, Sarah Pipes, Tyler Fischer, Jason Johnson for outstanding research assistance. In addition, we wish to thank Rachel Croson, Abigail Brown, Neslihan Uhler, Stefano DellaVigna, Ulrike Malmendier, Vera L. te Velde , David Danz, participants of the STIET and Berkeley Economics and Psychology seminar, ESA Tuscon 2010 conference and the International ESA conference 2011.

1. Introduction

Although very incomplete contracts are thought to perform substantially worse than their more complete counterparts, they are prevalent in the real world (Tirole 1999; Scott 2003) and, in many cases, lead to more efficient outcomes than more complete counterparts (Fehr and Falk 1999; Falk and Kosfeld 2004; Sliwka 2007; Rigdon 2009). Even incomplete contacts that take the form of unenforceable ‘cheap talk’ agreements, theoretically identical to having no contract at all, can and do substantially increase efficiency relative to the absence of a contract and relative to contracts that have some enforceable components (Charness and Dufwenberg 2006; Kessler and Leider *forthcoming*). As an example, in two of the contracting environments Kessler and Leider examine, the Double Dictator Game (a game in which two paired subjects are asked to transfer simultaneously their endowments to their counterparts) and the Bertrand Game (a game fashioned analogically to the Bertrand Competition model), informal agreement significantly shifts subject behavior towards the behavior prescribed by the informal agreement.

Explanations of why incomplete contracts do so well invariably rest on social norms such as reciprocity (Malhotra and Murnighan 2002; Dur et al. 2010; Englmeier and Leider 2010), fairness (Fehr and Falk 1999) or obligation and entitlement (Hart and Moore 2008) which are thought to motivate participants to act in socially beneficial ways that, in turn, enhance efficiency. As a result, social norms have received increased attention in the theoretical and empirical study of contracts.

A growing group of papers extend theory by allowing social norms to play an important role in determining ex-ante expectations about what actions ought to be taken or what actions will likely be taken by contracting parties (Sliwka 2007; Hart and Moore 2008; Fehr et al. 2009a). However, empirical work in economics generally tests for the presence of social norms such as inequity aversion (Fehr and Gachter 2000), entitlement (Fehr et al. 2009b) and reciprocity (Malhotra and Murnighan 2002; Dufwenberg and Kirchsteiger 2000; Seinena and Schramb 2006) by demonstrating behavior consistent with an influence of social norms.[2] Thus, as an example, in Kessler and Leider's experiments, they find that in the Double Dictator Game with no agreement, 14% take action "10" while 31% do so in the presence of an agreement. The authors posit that a social norm of honoring agreements is the mechanism that causes the shift in behavior toward the agreed upon action. Aside from inferring norms from the behavior, researchers have used other ways to identify social norms for a particular game; notably, these include directly asking subjects what one *ought* to do, eliciting beliefs about the most common action to be taken or asking what they would do in a particular context. But, if we ask subjects whether one *ought* to take action "10" in the aforementioned dictator scenario, 64% respond that one ought to take this action. However, only 10% of those asked believe that a majority of people (more than 50%) will actually take this action and only 28% say they themselves would choose to take action "10" themselves.

In this paper we depart from the previous literature by directly and separately identifying the norms that influence behavior, demonstrate how they vary across decision making environments and show that a utility model that includes social norms as an additional motivation does much better predicting behavior than models which include only monetary utility and social preferences for fairness. Specifically, we apply a new

[2] But see also Fehr et al. 1998.

method of identifying social norms using an economic experiment that is incentive compatible (Krupka and Weber 2008, Burks and Krupka forthcoming) to elicit social norms in two theoretically equivalent, but superficially dissimilar, choice environments: an environment where a 'handshake' agreement exists to take the first best action or an identical environment where no such agreement has been reached.[3] In addition, we examine the effect of these choice environments on norms for two different games, the Double Dictator and the Bertrand Games, where recent laboratory experiments by Kessler and Leider demonstrate that handshake agreements to take the first best action lead to substantially more pro-social choices and outcomes than identical decision making contexts where no such agreement exists (Kessler and Leider, *forthcoming*).[4] We then combine the results from our norm-elicitation experiments with the Kessler and Leider behavior data to run several horseraces between competing models. Our approach yields several distinct contributions.

By measuring social norms for these *two contracting environments* (the presence of a handshake agreement and the absence of such an agreement) we provide direct evidence of the central role that social norms play in affecting the choices Kessler and Leider observe. By measuring social norms for these *two games* (the Double Dictator and Bertrand Games), we obtain direct evidence of the common effects of handshake agreements on social norms across games.

We find that handshake agreements change the social norm profile in three ways. First, handshake agreements increase the appropriateness of the agreed-upon action, and

[3] This type of agreement can be thought of as a form of 'cheap talk' since the parties engage in 'costless', 'non-binding' and 'non-verifiable' messages (see Farrell and Rabin 1996). See also Charness and Dufwenberg (2006); Vanberg (2008) who look at unenforceable promises.
[4] For more on the Bertrand game see Dufwenberg and Gneezy (2000) and Dufwenberg et al. (2007).

decrease the appropriateness of all other actions. Second, handshake agreements decrease the variance between subjects in how appropriate the agreed-upon action is, i.e. they are better able to anticipate the group's perception of how appropriate actions are. Third, in the case of the Bertrand Game, the handshake agreement changes what action is seen as the most appropriate.

Third, we incorporate the elicited norms into our analysis to test whether they improve our predictive power; in doing so, we identify a distribution of norm-concern in the population that is consistent with a simple model where norm compliance is modeled as a source of utility. In our model the individual cares about both the payoff produced by the selected action and the degree to which the action is socially appropriate (as well as potentially social preferences for fairness). Using the model and behavioral data from the Kessler and Leider experiments, we show that the social norms we elicit significantly improve the predictive power of the model and capture key moments of the choice distribution (in particular the large fraction of subjects choosing the highest action under a handshake agreement). These results are important because they provide definitive evidence on the most prominent mechanism by which incomplete contracts are thought to enhance efficiency -- social norm compliance.

The balance of the paper is as follows. Section 2 provides a definition of social norms that motivates our method of eliciting norms using a coordination game experiment. Section 3 presents a simple model of utility that include social norms and outlines the predictions that come from the model. Section 4 describes the experimental design. Section 5 presents results and section 6 concludes with a discussion.

2. Defining and Identifying Social Norms

We define (injunctive) social norms as *jointly recognized beliefs, among members of a population, regarding the appropriateness of different behaviors.* Following Elster (1989), we note two important features of social norms. First, social norms generally prescribe or proscribe behaviors or actions, rather than outcomes. Allowing norms to govern actions, rather than outcomes, suggests that two actions that produce the same outcome, but differ in other respects, may be governed by different social norms (cf. Krupka and Weber 2008). Second, the "social" element of norms requires that they be jointly recognized, or collectively perceived, by members of a population.[5] These two features – that social norms typically apply to actions rather than outcomes and that they must be jointly recognized – are present in most researchers' definitions (Bettenhausen and Murnighan 1991; Fehr and Gächter 2000; Bicchieri 2006). For example, Ostrom (2000) defines social norms as "*shared understandings* about *actions* that are obligatory, permitted, or forbidden" (pp. 143-144, emphasis added).[6]

Further, we distinguish norms regarding what one "ought" to do, or injunctive norms, from customs or actions that people regularly take, or descriptive norms (Deutsch and Gerard 1955; Bicchieri 2006). Both kinds of norms influence behavior (Cialdini et al. 1990; Krupka and Weber 2009; Bicchieri and Xiao 2009). However, our focus here is on injunctive social norms, i.e., those described by Elster as prescribing what one "should do"

[5] At least implicitly, most definitions distinguish between social norms and personal norms. The former, which are our focus here, usually refer to a common understanding among members of a group. An individual member of a group has a belief that others in the group judge a particular behavior appropriate (or inappropriate) and that the others in the group assume the individual is aware of this judgment. In this sense, the individual and the group *share an understanding regarding the in/appropriateness of behavior* and this shared understanding is a social norm (cf. Bicchieri 2006; Young 2008).

[6] This is not to say that norms aren't also attached to outcomes, rather, these definitions give particular prominence to the actions associated with achieving outcomes. What we find in this paper is that if we maintain this simple assertion (that norms apply to actions rather than outcomes) we can already do much by way of identifying their role in decision making.

or "should not do." As we will show, social norms concerning the appropriateness of behavior one *ought to engage in* (injunctive norms) can explain a considerable amount of variation in behavior above and beyond the effect of subjects' beliefs about the descriptive norm.[7] From here on, when we talk about injunctive social norms, we will refer to them as social norms. When we wish to distinguish (injunctive) social norms from actions taken by most others, then we will refer to the latter as descriptive norms.

To measure the extent to which actions are jointly recognized to be socially appropriate or inappropriate we present respondents with a description of a choice environment, including all the possible available actions. We ask respondents to judge the social appropriateness of *each* action on a six point scale that ranges over "very socially inappropriate", "socially inappropriate", "somewhat socially inappropriate", "somewhat socially appropriate", "socially appropriate", and "very socially appropriate". In this sense, the technique is very similar to hypothetical vignettes used in psychology to identify social norms (some recent examples include Conroy and Emerson 2006, Ergeneli 2005, McKinney and Moore 2007, Gino et al. 2008, Oumlil and Balloun 2009). However, we provide respondents with incentives to match their ratings to the responses of other subjects in the session rather than to provide us with their personal opinions. Thus, respondents play a coordination game, with a proper scoring rule (Schelling 1960; Mehta et al. 1994) in which their goal is to anticipate the extent to which others will rate an action as socially appropriate or inappropriate, and to respond accordingly.

[7] In the experiment we isolate the influence of descriptive norms in two different ways that we describe in the design below.

Because social norms reflect "collective perceptions," coordination games present a useful incentivized way to identify such jointly recognized judgments.[8] From a game-theoretic point of view, matching games such as the one we use in our experiment have a number of equilibria, and nothing intrinsic to the game makes one equilibrium favored (or focal) over the other. Schelling (1960) theorized and Mehta et al. (1994) and Sugden (1995) demonstrated that prominence derived from common culture and shared experiences can create focal points.[9] In our experiment, we assume that collectively-recognized social norms create focal points in the matching game and we examine this assumption in the results. That is, if there is general social agreement that some actions are more or less socially appropriate, respondents attempting to match others' responses are likely to rely on such shared perceptions to help them do so. Thus, responses to our coordination game will not capture personal perceptions of the appropriateness of behaviors, but instead will capture collective perceptions of appropriateness, which we have defined to be social norms.[10]

3. Model and Predictions

Our instrumental goal is to use the elicited social norms from the coordination task to test for the effect of norms on behavior. To do so, we sketch a simple model to motivate our empirical work. We model the individual as caring about both the payoff $x_i(a_k)$

[8] Camerer and Fehr (2004) note that coordination games can be used with economic incentives to reveal shared understanding. They go on to suggest that experimental paradigms, such as simple coordination games, could prove useful for measuring dimensions of shared perception. See also Leider et al. (2009).
[9] Many previous researchers have noted the important relationship between social norms and equilibrium selection in games (Kandori 1991; Young 1998).
[10] Burks and Krupka (*forthcoming*) and Krupka et al. 2008 show that social norms elicited using the coordination exercise track ex-ante identified social norms and show that personal norms are distinct from those social norms. Krupka and Burks demonstrate the separate effect of these norms on behavior.

produced by the selected action, a_k, and the degree to which the action is compliant with norms for the particular reference group, g.[11]

$$u_i(x, a_k) = V_i\left(x_i(a_{i,k})\right) + \gamma_i N_g(a_k) \quad (1)$$

For an individual, i, the function *V( )* represents the value the individual places on the monetary payoffs from a particular action, a_k, and is concave and increasing in $x_i(a_{i,k})$.[12] One important feature of this model is that *actions* are arguments in the utility function; in this sense, the model departs from models that take social preferences into account, such as Fehr and Schmidt (1999), but can accommodate such preferences in the functional form of *V( )*. By doing so, we allow for perceived appropriateness to come through on the actions rather than the outcomes without diminishing the importance of outcomes to utility nor diminishing the trade-off between willingness to sacrifice to uphold a norm and achieving other personally valued instrumental goals.[13]

We let $A=\{a_1,\ldots,a_K)$ represent a set of K actions available to a decision maker. The social norm function $N_g(a_k)$ assigns to each action a degree of appropriateness or inappropriateness that reflects the norm of the relevant reference group, g. Thus if, for an action, a_k, there is collective recognition among group members that the action constitutes "norm consistent" behavior for that group then $N_g(a_k) > 0$. If there is joint recognition that an action constitutes "norm inconsistent" behavior for that group then $N_g(a_k) < 0$. The definition of a social norm applies to the *entire set of possible actions* and,

[11] Here the reference group is defined as fellow subjects participating in the particular session.

[12] In strategic interactions, the payoffs are also a function of other's actions. However, for simplicity of exposition, we have suppressed this argument in the utility function.

[13] In most cases, taking actions consistent with a social norm requires that actors have a personal commitment to the norm that cannot be captured by the self-regarding "public payoffs" (eg. avoiding punishment or garnering praise) associated with complying with the norm. One explanation for the theoretical micro foundations of such a commitment is that it arises from evolved psychological predispositions that render norms effective (Masclet et al. 2003; Andreoni and Bernheim 2009).

as such, the social norm function characterizes a *profile* of appropriateness ratings over all the actions available to a decision maker.[14]

The parameter γ_i represents the degree to which the individual cares about adhering to a particular group norm. Several researchers have noted that there exists heterogeneity among individuals for the degree to which they care about complying with a social norm (cf. Ostrom 2000, Fisher and Huddart 2008).[15] An individual entirely unconcerned with social norms ($\gamma = 0$) will always select the payoff-maximizing action. On the other hand, as γ_i increases, an individual will derive greater utility from selecting actions that are socially appropriate relative to the utility from those that are not.[16]

While the above formal definition and model are straightforward, they present a useful framework for understanding how behavior might change across choice environments even when they are payoff-equivalent. It also provides a testable relationship between the degree of social appropriateness of actions and individuals' willingness to take those actions, provided one has a reasonable method for capturing the "social appropriateness" of the different available actions. Our hypotheses articulate how appropriateness ratings elicited in our experiment change in the presence of agreement. Subsequent tests, in section 5.3, examine the predictive power of social norms for various assumptions about the underlying utility model we articulated in (1).

---

[14] That is, a norm is not necessarily a binary classification, such that a particular action (the "norm", e.g., "tip 20%" or "the 50-50 split") should be taken, by assumption leaving all remaining actions as those (equally inappropriate) actions that should not be taken. Such a definition is possible in our framework (by for example, assigning $N(a_k) > 0$ to only one action (the "norm") and letting all other actions have a constant value of $N(a_k) < 0$) but is an over simplification of how norms appear to operate.

[15] Such heterogeneity in pro-social concern is also common in most models of social preferences (Fehr and Schmidt 1999; Andreoni and Miller 2002; Benabou and Tirole 2006).

[16] Cases in which $\gamma_i < 0$, which we do not explore here, might correspond to individuals who are anti-social, or derive utility from violating norms.

In previous research Krupka and Weber (2008) find that subjects judge pro-social behavior as generally socially appropriate while more selfish behavior is generally considered less socially appropriate (though the relationship is not clearly monotonic). In the context of our experiment, this leads to the following straightforward hypothesis regarding the appropriateness ratings:

> *Hypothesis 1: Actions that are more prosocial will be seen as socially appropriate, and actions that are more selfish will be considered less socially appropriate.*

Numerous experiments have demonstrated that pre-play communication of various forms can increase the prosociality of individual behavior (see for example Dawes et al. 1977 for an early experiment, and Sally 1995 for an early survey). Promises to take a particular action have been shown to be particularly powerful in changing behavior (Charness and Dufwenberg 2006, Vanberg 2008, Kessler and Leider 2010). These results can be interpreted to suggest that there is a norm of promise-keeping that will be active in the Handshake treatment. For example, while sending 8 tokens in the Double Dictator Game may be seen as relatively prosocial when there is no agreement, it is a violation of the promise to send 8 tokens in the same game where subjects have a handshake agreement to send 10 tokens. Thus, sending 8 in the former case may be judged "socially appropriate" while sending 8 in the latter case my judged a "socially inappropriate" action. Thus, handshake agreements to take a particular action, should make the promised action seem more appropriate (compared to cases where no such agreement has been discussed), and the other actions as less appropriate. This yields hypothesis 2 and 3:

> *Hypothesis 2: When participants have agreed upon a particular action, taking that action will be seen as highly appropriate, and*

any action not consistent with that agreement will be seen as substantially less appropriate.

Hypothesis 3: When participants have agreed upon a particular action, then the variance in appropriateness ratings will decrease around that agreed upon action relative to the variance around that action in the No Agreement treatment.

4. The Experimental Design

The experimental design consists of five modules.[17] The first and second module elicit the two different kinds of norms (injunctive and descriptive) and, together, comprise the primary data we use to investigate our research questions. The remaining three modules that follow provide data with which we perform a series of robustness checks on our primary findings and claims. Regardless of the treatment condition, all subjects participate in all modules and the order in which subjects see the modules is always the same. In all cases, subjects are informed of their individual earnings only after all experimental modules were completed. We now describe each of these modules in detail.

4.1 Eliciting injunctive norms using coordination games (Module 1)

The first module, “Injunctive Norms, Initial”, uses coordination games to elicit subjects’ beliefs about normative evaluations, and in aggregate, identifies the social norm for that decision context. In our experiments we elicit the social norms for the same choice environments that Kessler and Leider (2010) use because this will allow us to make useful comparisons between the actual behavior of subjects in the Kessler and

[17] See Table S1 in the Supporting Materials for an overview of the experimental design.

Leider experiments and our norm ratings from the current experiment. Thus, in the Double Dictator Game with no agreement subjects read the following vignette[18]:

> *Individual A and Individual B are randomly paired with each other. This means that A and B do not know each other and will never find out who the other person is. A and B each start with tokens worth 20 units. A must choose an action. B will also be choosing an action at the same time. The action that A and B choose will determine their earnings. A and B are told that their payoffs will be calculated in the following way: A's earnings are 20 - (2 × what A sends) + (6 × what B sends). B's earnings are: 20 - (2 × what B sends) + (6 × what A sends). Beyond these basic instructions, [*in the case of No Agreement] *A and B were not given the opportunity to make any kind of agreement about what action they were each going to take.*

Subjects reading about the Bertrand Game with no agreement read the following vignette:

> *Individual A and Individual B are randomly paired with each other. This means that A and B do not know each other and will never find out who the other person is. A must choose an action. B will also be choosing an action at the same time. A's action, and B's action, can be to select any whole number between 0 and 100. Whoever chooses a smaller action has a payoff equal to his action while the other player gets a payoff of zero. If A and B choose the same number, then their payoff will be equal to ½ of that number. [*in the case of No Agreement*] A and B were not given the opportunity to make any kind of agreement about what action they were each going to take.*

In the Agreement treatments, subjects were instead told that "A and B were given the opportunity to make an agreement about what action they were ach going to take. They agreed to each take action 10 [100]".

[18] Both vignettes are abbreviated for exposition purposes. The entire set of instructions is available and can be found in the Supporting Materials II.

After reading about the situation and completing a comprehension check[19], subjects are asked to evaluate the social appropriateness of each of the actions available to A and to rate how sure they are that each of their ratings will match with each of the ratings of another subject. Subjects only rated one game (either the Double Dictator Game or the Bertrand Game) for one contracting environment (either with Agreement or No Agreement).

Figure 1 depicts the decision screen subjects saw for the Double Dictator Game. We ask respondents to judge the social appropriateness of *each* action on a six point scale that ranges over "very socially inappropriate" to "very socially appropriate."

Figure 1. Screenshot

We tell subjects that by "socially inappropriate" we mean "consistent with what most people expect individual A ought to do". We also tell them that we will pay them not to reveal their own personal preferences but instead to try and match the appropriateness ratings of others. To incent subjects to think about what most others think is appropriate, we use a proper scoring rule (Lambert and Shoham 2009). This scoring rule elicits subjects' median belief[20] about the distribution of others' ratings by matching a subject with another subject and then paying them according to the following payoff function:

$$u_i = 1.5 - 0.4(x_i - x_{-i}), \text{ for each subject } i \quad (2)$$

[19] Subjects were also tested on their comprehension of the situation with an interactive quiz, in which they calculated the payoffs of both players in three hypothetical situations. They were not allowed to proceed until they got all the calculations correct.

[20] 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 u_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.[21] In order to test our hypotheses, we converted subjects' norm ratings into numerical scores. A rating of "very socially inappropriate" received a score of 1, "socially inappropriate" a score of 2, "somewhat socially inappropriate" a score of 3, "somewhat socially appropriate" a score of 4, "socially appropriate" a score of 5 and "very socially appropriate" a score of 6.[22]

In addition to the coordination exercise, subjects were asked about how sure they were of their ratings for each action. The sureness ratings were not incentivized and were on a 4 point scale: "I am not sure", "I am somewhat sure", "I am fairly sure" and "I am very unsure" that my appropriateness rating will match that of another person.

4.2 Eliciting descriptive norms (Module 2) and observing behavior (Module 3)

After subjects complete the "Injunctive Norms, initial" module, a second "Descriptive Norms" module asks our subjects to tell us how they think the subjects in the original Kessler and Leider experiment actually played the game. Specifically, our subjects told were asked to guess the *modal action* for the choice environment, the *percent who took the highest action* and the *percent who took the lowest action*. We

[21]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. 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 (0.4) 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.

[22]In so doing we are imposing ratio scale characteristics on measurements that are in design ordinal. In some of what follows this is merely for convenience, such as when we use a rank-order test for the equality of distributions. But on other occasions it implicitly adds extra assumptions upon which our analysis is then conditional, such as when we compare means.

incentivize responses using the same proper scoring rule as in the 'Injunctive Norm, initial' elicitation module:

$$u_i = 1.5 - 0.4(x_i - c), \text{ for each subject } i \tag{3}$$

Here u_i is the payoff of subject i, and x_i is the subject's guess for the question, and c is the correct answer to this question derived from empirical data from Kessler and Leider's experiment.

We then show our subjects the choices of five of the Kessler and Leider subjects. There are three between-subject variations on what subjects are shown. The five decisions are either (1) drawn randomly from the entire empirical distribution of choices, (2) drawn randomly from the distribution but with an upper tail bias or (3) drawn randomly from the distribution but with a lower tail bias.

In the "upper tail biased" draw ("lower tail biased"), the five observations contain three observations from the upper tail (lower tail) of the distribution, one from the middle and one from the lower tail (upper tail) of the distribution.[23] Thus, depending on which 'observation condition' subjects were in, they were exposed to a different descriptive norm – where a majority of decisions come from the upper or lower portion of the action space.[24] We can use this module to examine the effect of observing a descriptive norm on subject's perception of the injunctive norm by having subjects repeat the injunctive norm elicitation task again in Module 4.

[23] See Table S4 to S7 of the Supporting Materials for the distribution of actual choices observed by our subjects.
[24] Subjects were told that they would observe 5 randomly drawn observations though no details were provided with respect to the skew of the draws.

4.3 Second elicitation of injunctive norms using coordination games (Module 4)

In our fourth module, “Injunctive Norm, after”, subjects were asked to complete the injunctive norm rating again for the same vignette they rated in the “Injunctive Norm, initial” module. The task, procedures and the incentive scheme were the same as for Module 1. Because subjects were exposed to different (biased) sets of observations in Module 3, Module 4 allows us to assess whether a subject’s beliefs about the injunctive norm change after observing the behavior of individuals in that context.

4.4 Collecting control variables of interest (Module 5)

Lastly, we ask subjects to make choices in two games that may give us proxy measures of their willingness to adhere to social norms. In the "Advice Game" (Gneezy, 2005), we obtain a direct measure of the willingness to be honest at a significant financial cost while controlling for first mover beliefs about the likely responses of second movers. Each participant is anonymously paired with a counterpart for a one-time decision. There are two options: Option A pays \$10 to the first mover and \$5 to the second mover, while Option B pays the reverse amount. The first mover’s only action is to send a message to the second mover that a particular action will give the second mover a higher payoff. The second mover’s only action is to decide which option is implemented but he is not told the payoffs associated with the options.

We also measure subjects’ willingness to assist another with the “Helping Game”. Subjects are randomly and anonymously re-matched into a pair, and each of them is assigned a different role. One member of the pair is in the helping role and has \$12 while the other has \$0. Subjects in the helping role can increase their matched participants’ earnings by \$6 if they pay a price \$P, where the amount \$P is drawn randomly between

$0 and $6. Subjects in the helping role state the highest amount of $P that they are willing to pay, the computer randomly draws the price and determines whether the price is above the stated willingness to pay. Taken together, the Advice Game and the Helping Game can be thought of as proxy measures of a subject's willingness to comply with a proscriptive ("do not lie") social norm and a prescriptive ("do help others") norm. The experiment concludes after our last module, in which subjects fill out a demographic questionnaire that asks about gender, age, race and ethnicity, and religious affiliation.

4.5 Payment procedures

Subjects' earnings for the experiment were calculated using the coordination payoffs described above. Their payment was calculated from three different components: (1) from one randomly selected action rating *either* from the "Injunctive Norm, initial" module or the "Injunctive Norm, after" module, (2) their payoffs from the guesses about behavior in Module 2 and (3) their payoff from *either* the Advice Game or the Helping Game in Module 5 (randomly selected). Subjects also received a $5 show-up fee. Subjects were paid privately at the end of the experiment.

5. Results

Students from the University of Michigan were recruited to take part in our experiment, and a total of 356 participants were recruited in 36 sessions. Sessions were conducted using an even number of participants, ranging from 6 to 22 per session and the average length of each session was one hour and fifteen minutes. All experimental instructions were read aloud and shown on a screen. The average payoff for each subject was $29.72. Table 1 details participation and average payoffs by treatment.

Table 1 about here.
(Experimental Overview)

5.1 Injunctive norm ratings with and without agreement across both games

Figure 2 displays the average appropriateness ratings for the Double Dictator Game with and without a handshake agreement from the Injunctive Norm Elicitation Module 1. In both treatments sending a small amount is seen as fairly socially inappropriate, while sending a large amount is seen as quite appropriate. This suggests that subjects are using the full range of appropriateness ratings and that their ratings are consistent with Hypothesis 1. That is, actions which are generally more prosocial are considered appropriate and actions that are more selfish are considered less socially appropriate.

Figure 2 about here:
(Average appropriateness ratings for DDG)

However, there are notable differences between the environment where an agreement exists and where none exists. First, every action other than sending the full amount is seen as less appropriate in the Agreement treatment than in the No Agreement treatment. A rank-sum test finds that appropriateness ratings are significantly higher in the No Agreement treatment than in the Agreement treatment for actions 0 to 9 ($p < 0.01$ for all). Second, sending the entire endowment of ten tokens is seen as more appropriate in the Agreement treatment than in the No Agreement treatment ($p < 0.01$). Additionally, the greatest increase in appropriateness in the No Agreement treatment is for relatively low actions, while the ratings change little and remain fairly flat for high transfer decisions; in particular the average rating for sending all ten tokens is not significantly higher than sending all nine tokens (signed-rank test: $p = 0.52$). By contrast, there is a very large difference in the Agreement treatment where sending as much as nine tokens in the Agreement treatment is rated as being roughly neutral but sending all ten tokens is rated as being very appropriate (signed-rank test: $p < 0.01$).

Figure 3: about here.
(Average appropriateness ratings for BG)

In Figure 3 we plot the average appropriateness ratings for the Bertrand game for both the Agreement and No Agreement treatment. As in the Double Dictator Game, choosing a small action is seen as fairly socially inappropriate while choosing a large action is seen as appropriate in both the Agreement and No Agreement treatments. A rank-sum test supports Hypothesis 1: for actions less than 100, average ratings in the No Agreement are greater than average ratings in the Agreement ($p=0.03$ for action 0, $p < 0.01$ for all others). Choosing action 100 is considered more appropriate in the Agreement treatment than in the No Agreement treatment ($p<0.01$). In fact, in the No Agreement treatment the average appropriateness rating increases from 0 to 50, peaks at the middle (action 50) and declines from 50 to 99 (signed-rank test of appropriateness ratings in the No Agreement treatment for 'action 50' > 'action 40' is $p < 0.01$; 'action 50' > 'action 60' is $p=0.04$). Moreover, in the No Agreement treatment there is no significant difference in appropriateness rating between 'action 50' and 'action 100' (signed-rank test: $p = 0.57$), while in the Agreement treatment choosing 'action 100' is significantly more appropriate than any other action ($p < 0.01$ for all comparisons). In summary, these graphs reveal the strong effect of agreement on the entire profile of appropriateness ratings. In the presence of an agreement, all actions except the agreed upon action are considered substantially less appropriate than (a) the agreed upon action and (b) the same actions when no such agreement has been reached.

Table 2 about here.
(Main regression table)

Regression analysis supports these results and highlights a few key findings. We begin with the simplest specification, presented in columns 1 and 4 of Table 2; these columns report the results of regressing for each action the subjects' appropriateness rating on the action and a dummy for the agreement treatment. This captures the simplest forms of Hypotheses 1 and 2, that more prosocial (higher) actions are deemed more appropriate, and that high actions should be particularly appropriate in the Agreement treatment. This specification does a reasonable job of capturing the patterns we saw in Figure 2 in the Double Dictator Game – there is a positive coefficient on the variable 'action' (b=0.275, $p < 0.01$) and the increase in appropriateness for higher actions becomes steeper in the Agreement treatment (b=0.078, $p < 0.01$).

However this specification is not flexible enough to capture the non-monotonicity and the sharp discontinuities in the Bertrand Game very well. The positive effect on the appropriateness of 'action 100' (b=0.0758, $p < 0.01$) is swamped by the overall negative effect on all other actions (b=-1.106, $p < 0.01$). To that end, in the specifications reported in columns 2 and 5 of Table 2, we add an additional dummy variable denoting the 'highest action'. This allows us to capture the "jump" in ratings at the highest action in the Agreement treatment, and aligns with our a priori prediction (hypothesis 2) that handshake agreements should change the perception of the promised action. In both specifications it is clear that there is a substantial increase (b=2.298, $p < 0.01$ in the dictator game and b=3.423, $p < 0.01$ in the Bertrand game) between the highest and next highest actions in the Agreement treatment – an increase not matched in the No Agreement treatment. Furthermore, in the Bertrand Game the net effect of the estimated coefficients in the Agreement treatment is that the appropriateness ratings should be flat for all actions less than 100, with a sharp increase at 100. In short, the regressions restate what the graphs show: the effect of a handshake agreement is to increase the appropriateness of the agreed-upon action, and decrease the appropriateness of all other actions.

One advantage to using the norm elicitation technique is that it allows us to identify ex-post other important changes in appropriateness ratings that we wouldn't have known about before eliciting the norms. In Figure 4 we see a peak in appropriateness for 'action 50' in the Bertrand Game. Based on this observation, we add an additional dummy for the 'middle action' in Columns 3 and 6 of Table 2. In these specifications we can see that the 'middle action' is rated as slightly above the trend in the No Agreement Double Dictator Game, and substantially above trend in the No Agreement Bertrand Game, but that both of these differences are eliminated in the Agreement treatment. Though not anticipated ex-ante, the norm elicitation procedure leads to the additional insight that, in the case of the Bertrand game, the handshake agreement changes *which* action is seen as the most appropriate rather than simply making a particular action *more* appropriate (as in the Double Dictator game). Put differently, these results suggest that agreements have *the power to change* the social norms profile – a result we interpret as a change in the norm.

Qualitative data collected at the end of our session supports the interpretation that social norms fundamentally change when there exists an agreement in the Bertrand Game. In the situation where there is no agreement, subjects appear to rely on a norm of risk minimization but rely on a norm of honoring obligation when an agreement has been reached. Thus, as an example, a subject in the No Agreement condition stated the following reason for his appropriateness ratings: “I felt that the higher options were more inappropriate because of the risk factor so the ones in the middle were somewhat appropriate, and the ones that were from 0-10 were inappropriate because you would win little or no money”. However, when there was an agreement subjects used different language to describe the thoughts that guided their appropriateness ratings. As an example, a subject in our Agreement condition wrote: “I thought that any action that violated A and B's agreement was socially inappropriate regardless of A's resulting pay off.”

Table 3 about here.

We analyze the effect of the Agreement treatment on the shape of the appropriateness profile more formally in Table 3. Here we construct two summary measures for the jump in ratings for the highest action, and for the middle action. Specifically, we define the “top jump” for each subject as the difference in their rating for the highest action and the second-highest action, and we define the “middle jump” as the difference between their rating for the middle action and the average rating for the two neighboring actions. In columns 1 and 2 of Table 3 we see that there is an increase in the ratings difference in the Double Dictator Game of 1.9 appropriateness categories in the presence of a handshake agreement, and an increase in the difference of 2.7

appropriateness categories in the Bertrand Game with an agreement. In columns 3 and 4 we see that there is no difference between treatments in the shape of the ratings around the middle action in the Double Dictator Game (column 3), while in the Bertrand Game the peak at 50 exists only in the No Agreement Treatment (column 4). Taken together with the qualitative data, this evidence is consistent with an interpretation that subjects seem to apply different social norms in the Bertrand game when there is an agreement and where there is not.

5.2 Variance in Norm Ratings with and without agreement

We now turn our attention to examining how the presence of an agreement can affect the variance in the subjects' judgments of an action's appropriateness. We can think of the variance in appropriateness ratings as a measure of the concentration of beliefs about social norms. It would be a straightforward extension of the norms model to assume that the weight of social norms (relative to monetary payoffs) in an individual's utility is increasing with the level of concentration in others' beliefs in the appropriateness of each action. Figures 4 and 5 display the standard errors around the average rating for each action in each of our four treatments.

Figure 4a-4b: Norm ratings and standard errors for DDG
Figure 5a-5b: Norm ratings and standard errors for BG

In both games we find that the variance in ratings increase dramatically for higher actions in the No Agreement treatment – that is, subjects on average think higher actions are more appropriate, but they are poorly coordinated on how much more appropriate (ie, what the utility payoff will be from taking these actions).[25] In particular, the rating for sending 10 tokens in the Double Dictator Game has a higher variance in the No Agreement treatment than each of the ratings for sending 8 or fewer tokens (robust variance test: $p < 0.01$ for all), and similarly the rating for action 100 in the Bertrand Game has a higher variance than the ratings for any action of 96 or less ($p < 0.01$ for all).

In the Agreement treatment, however, for the Double Dictator Game we find much more variance in the ratings about the appropriateness of intermediate actions, but less variance about the appropriateness of sending 10 tokens. In particular, in the Double Dictator Game we find that there is significantly less disagreement about the appropriateness of sending 10 tokens in the Agreement treatment than in the No Agreement treatment ($p < 0.01$), and similarly that there is significantly less disagreement about the appropriateness of choosing action 100 in the Bertrand Game in the Agreement treatment than in the No Agreement treatment ($p < 0.01$). These results are consistent with hypothesis 3.

We can demonstrate this result another way by looking at how many subjects rate taking action '10' in the Double Dictator Game and taking '100' in the Bertrand Game as the highest rated (most appropriate) action. For the Double Dictator Game 64 percent say taking action '10' is their highest rated action when there is 'no agreement' but 94 percent say so when there is a handshake agreement and in the Bertrand Game the percentage is 57 and 90 percent respectively for taking action '100' (test of proportions: $p < 0.01$ for both). Similarly, we can examine how many subjects rate taking action '10' in

[25] We also examine the distribution of norm ratings using box plots (see Figure S1 – S4 in Supporting Materials). When there is no agreement, the ratings for the highest actions have one of the largest interquartile ranges among all of the actions. However, when there is an agreement, there is substantially higher coordination in the ratings for the highest action, with ratings having an IQR of 0 for both games.

the Double Dictator Game' and taking '100' in the Bertrand Game as the uniquely highest action (that is, no other action receives a higher or equal appropriateness rating). Here we find that in the Double Dictator Game this is 19 percent when there is no agreement and 76 when there is an agreement, and in the Bertand Game this is 21 percent and 83 percent respectively (test of proportions: $p < 0.01$ for both). These findings clearly demonstrate that handshake agreements decrease the variance between subjects in how appropriate the agreed-upon action is allowing them to coordinate more effectively in evaluating the appropriateness of the action.

5.3 Summary of Robustness Checks

In the Supporting Materials we also present several robustness checks of our main results. Here we summarize these findings briefly. In order for handshake agreements to have general and consistent effects on behavior we would like these agreements to evoke ratings in the coordination task that don't rely upon individuals who either have particularly optimistic beliefs about actions people are likely to take or upon particularly norm-compliant individuals. Furthermore, while social norms regarding what one ought to do provide focal points that subjects can rely on in the coordination games, beliefs about the descriptive norm (what people actually do) might also serve as focal points. That is, a change in the norms ratings across treatments could be consistent with either subjects using beliefs about the injunctive or the descriptive social norm to suggest focal points. Lastly, it may be that the effect we claim to observe of handshake agreements on social norms is strongly determined by the fact that subjects agree on the pareto efficient action in the handshake agreement. We examine these concerns in the appendix and summarize the results here.

To explore the effect of beliefs on norm ratings, for each game and each treatment condition, we divide subjects into two categories based on their stated belief (obtained in module 2) for the modal action taken in the original Kessler and Leider experiment: those whose beliefs are above the median belief or below the median belief for their treatment. We then calculate the average appropriateness ratings for each action for those two groups (Figures S5 and S6 in the Supporting Materials report these average ratings and Table S3 reports regression results). We find that subjects with more optimistic beliefs tend to have lower ratings for low actions, and higher ratings for high actions. However, it is clear that the major difference in the shape of the norm function between treatments is not driven by beliefs.

Injunctive norm measures obtained in Module 1 are also not different for those subjects who might be characterized as relatively norm-compliant. Using their choices in the 'Advice Game' and the 'Helping Game' we can create proxy measures of individual norm-compliance. First movers in the Advice Game who send an honest message and believe that the second mover will follow their advice are typed "Honest", those who send an honest message but do not believe that the second mover will follow their advice are typed "strategic honest", those who send a message that is a lie and believe that the second mover will take their advice are typed as "liars", and those who send a lying message but who do not believe that the second mover will take their advice are typed "altruistic liars". Thus, actions coupled with beliefs by the first mover in the Advice Game allow us to characterize subjects who have some willingness to adhere to social norms against lying (those are the "altruistic liars" and "honest" subjects) and those who have a lower willingness to adhere to social norms against lying (those are the "strategic honest" and "liar" subjects). For the Helping Game, we characterize those helpers whose willingness to pay is among the top range of all helpers as the "high helpers", those subjects whose willingness to pay is among the middle range of all helpers as the "middle helpers", and those whose willingness to pay is among the bottom range of all helpers as the "low helpers". Choices in the Helping Game allow us to characterize subjects who have some willingness to adhere to a prescriptive norm of helping others. One might expect that if subjects are using their own preferences (and implied appropriateness ratings) as focal points in the coordinating task, then subjects who care about pro- and prescriptive norms (both or just one) may provide different injunctive norms ratings than those who care less. However, we find no difference in the injunctive norms ratings along these personal characteristics or demographics (see Table S11 and Table S12 in the Supporting Materials).

In order for handshake agreements to have a persistent effect on behavior we would like them to be resilient to observing an occasional failure to comply with the social norm. Recall, that our subjects observe 5 randomly drawn actions from what other subjects actually did in Kessler and Leider's experiment (2010). After observing what others actually did, our subjects perform the injunctive norm rating task from module 1 again. In our analysis of the effects of observation on norm ratings, we divided our subjects up into whether they observed draws that were "upper tail biased", "lower tail biased" or "random from the whole distribution". Using regression analysis, we find that different observations have no significant effect on appropriateness ratings in the second injunctive norm elicitation stage (p=0.264, p=0.621 for 'high-' and 'low-observation' dummy variables in the Double Dictator Game and p=0.243, p=0.432 in the Bertrand

Game Table S8 respectively). We also do not find that injunctive norms elicited in module 1 (prior to guessing what others do or seeing what others did) differ from the injunctive norms elicited in module 4 (see Table S9 and S10).

Finally, we ran two additional sessions for the Double Dictator Game and the Bertrand Game to test the effect of a handshake agreement on the social norms profile when subjects are told that the agreement is to take an action *other* than the Pareto efficient one. In the Double Dictator Game we elicited social norms profiles from subjects were told that the agreement was to transfer 7 tokens and in the Bertrand Game the agreement was to take action 70. We find that agreeing on 'action 7' makes it significantly more appropriate to take 'action 7' than when there is no agreement ($p < 0.001$) and we find no difference in the appropriateness ratings for taking 'action 7' when there is an agreement on action 7 and 'action 10' where no agreement has been reached ($p = 0.949$). Our findings are similar (and more extensively described in the Supporting Materials) for the Bertrand Game.

These additional results are important because they bolster our claim that the norms we elicit are not sensitive to the type of person we obtain them from, to other reasonable focal points (beliefs about or observation of others' behavior), nor are they reliant on agreements over actions that are also the social optimum. We now turn to examining the predictive power of social norms for various assumptions about the underlying utility model.

5.4 Predicting Choice Behavior using Social Norms

In order to examine whether our measured norms can explain behavior in games, we fit individual utility functions to the choice data from Kessler and Leider (2010). Recall, that if norms are an important motivation for behavior, then a model that incorporates concern for norms ought to outperform models that do not. In the Kessler and Leider experiments, a subject made choices for both contracting environments and for both games. If we estimate the norm function N( ) from equation (1) by using the average norm rating elicited in our experiment for each action[26] then we can estimate the norm sensitivity γ, from equation (1), that best rationalizes subject choices in the Kessler and Leider experiments using maximum likelihood. 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:

$$P(a = a_i) = \frac{\exp(U_i)}{\sum_j \exp(U_j)} \tag{4}$$

We therefore estimate conditional logit regressions to compare how well our predicted frequencies matched up with the actual Kessler and Leider behavior when we assume the following:

1) that actors only derive utility from payoffs by setting $\gamma_i = 0$. We call this the '*selfish model'*.

$$u_i(x_, a_k) = x_i(a_{i,k})$$

2) that actors derive utility from payoffs *and* norms; we restrict gamma to be the same for everyone ($\gamma_i = \gamma > 0$). $N_g(a_k)$ is measured using the appropriateness ratings collected from our experiment. We call this the *'norms model'*

$$u_i(x_, a_k) = x_i(a_{i,k}) + \gamma_i N_g(a_k)$$

[26] For the Bertrand Game, we use linear interpolation to determine the appropriateness of the actions that we did not explicitly measure. The programs that produce these interpolations are available.

3) that actors have only *Fehr and Schmidt* preferences (1999).[27] We call this the *'FS model'*.[28]

$$u_i(x, a_k) = x_i(a_{i,k}) - \alpha_i \max\{x_j(a_{j,k}) - x_i(a_{i,k}), 0\} - \beta_i \max\{x_i(a_{i,k}) - x_j(a_{j,k}), 0\};$$
$$i \neq j.$$

4) Or that actors have both Fehr and Schmidt preferences and a desire to comply with the social norm. We call this the *'FS + norms' model*'

$$u_i(x, a_k) = x_i(a_{i,k}) - \alpha_i \max\{x_j(a_{j,k}) - x_i(a_{i,k}), 0\} - \beta_i \max\{x_i(a_{i,k}) - x_j(a_{j,k}), 0\}$$
$$+ \gamma_i N_g(a_k);$$
$$\gamma_i = \gamma > 0 \ \ and\,; \ \ i \neq j$$

Table 3 reports the regressions resulting from this exercise. Columns (1) and (5) report for the DDG and BG respectively the results of the 'selfish' model where subjects only care about their own monetary payoff. Specifications (2) and (6) include our norm measure as an additional component of utility and estimate the 'norms model' while columns (3) and (7) estimate the 'FS model'[29]. Lastly, columns (4) and (8) report a combined specification with both the norm ratings and the inequity measure ('FS+norms model'). Because the average norm rating was derived from our earlier experimental sample, and therefore contains sampling error relative to the true population norm, we construct bootstrapped standard errors (reported in brackets) by resampling subject norm profiles from the Module 1 data.

---

[27] Fehr and Schmidt (1999) characterize actors as having a weak urge to reduce inequality when on top and strong urge to reduce inequality when on the bottom. In particular, it is standard to assume $0 \leq \beta_i \leq 1$ and $\beta_i \leq \alpha_i$ reflecting the fact that people are more sensitive to disadvantageous inequality.

[28] We cannot separately identify coefficients for material payoff, positive inequity and negative inequity in the Double Dictator Game. We therefore fix $\alpha = 2$ and $\beta = 0.6$ (see Fehr and Schmidt 2004, Fehr et al. 2007 and Fehr et al. 2008) to construct the combined disutility of inequity, and estimate the relative weight subjects place on this utility component. We find very similar results if we use alternate inequity parameters (from Fehr and Schmidt 1999 we tested (0.5, 0.25), (1.0, 0.6) and (4.0, 0.6)).

[29] We use inequity aversion as an example of prosocial utility models. We find similar results from using conditional cooperation instead (i.e. disutility from taking a different action than the other subject).

In each regression, the reported coefficient reflects the relative weight that each component has in the utility function. We find that all three utility components play a significant role in subjects' choices. Furthermore, adding any component significantly improves the model's predictive fit, with the combined specification being the best fit ($p < 0.001$ in a likelihood ratio test for all comparisons of nested models).[30] Similarly, the Bayesian Information Criterion (which penalizes models for the number of parameters) improves with the additional components in both games. While the 'FS model' specification does particularly well in the Double Dictator Game, this appears to be driven by the fact that it is the only utility component in that game that includes information from the subjects' guesses. In the Bertrand Game that information is already contained in the monetary payoff term, hence adding inequity aversion adds less to the model. Graphing the actual data and the predicted distribution of choices for each model (displayed in Figures 6 and 7) visually attests to the relative importance of norms for explaining behavior. The results of our analysis tell a clear story: norm ratings are performing the heavy lifting of fitting key moments of the empirical distribution – in particular the mass of subjects choosing the highest action when there is an agreement and the mass of subjects choosing the middle action in the no agreement condition. Therefore, it is clear that in order to capture the effect of promises on subjects' actions, it is essential to include information on the norms that govern subject behavior.

6. Conclusion

[30] To compare the Norms and FS models, we use the test from Vuong (1989) for non-nested models. For both games the model with the lower BIC is significantly preferred ($p < 0.01$ for both).

Theory gives social norms a leading role to explain both the persistence and success of incomplete contracts. Empirical tests of these theories, however, do not identify the norms ex-ante, but identify observed behavior consistent with social norms. In this paper we directly elicit social norms and analyze their role in two different games and two different contracting environments. Our results provide direct evidence of the central role that social norms play in affecting choices in the presence of handshake agreements. First, taking the promised action becomes substantially more appropriate, and all other actions become less appropriate. Second, the handshake agreement increases the consensus across individuals about which action is the most appropriate. Third, in the Bertrand Game the handshake agreement replaces a norm of risk minimization with a norm of obligation. Finally, using the model and behavioral data from the Kessler and Leider experiments, we show that the social norms we elicit significantly improve our predictive power and capture key moments of the choice distribution. These results are important because they provide definitive evidence on the most prominent mechanism by which incomplete contracts are thought to enhance efficiency -- social norm compliance.

The evidence suggests at least three channels by which the act of making an agreement seems to operate on *behavior*: agreement makes a particular norm of obligation salient, it reduces confusion about what action one should take to satisfy that norm, and increases the utility cost of deviating from the obligation. In particular, an agent with a given level of concern for following the norm (γ in our model) will be more likely to choose an action consistent with the social norm in an agreement condition because he has an increased awareness that a social norm applies in this environment and because he is more sure about which action he should take to behave in a manner consistent with the social norm.

This work also offers compelling new findings regarding how norms vary from environment to environment that can allow for more detailed models. In particular, our results in the Bertrand Game suggest that strategic complements strongly affect the importance of complying with an agreement - any action that does not honor that agreement is rated as very socially unacceptable. No such dramatic shift in appropriateness exists when actions are strategically independent and an agreement has been reached.

Finally, our work advances a new methodology for identifying social norms. We show that using incentivized coordination games can elicit shared notions of appropriateness that are consistent with a social norms interpretation and that improve predictive power. A strength of this approach is that one need not know the particular social norm (is it fairness?, obligation?, reciprocity?) ex-ante, but can use this technique to characterize the social norm and make and test predictions about behavior.

References:

Andreoni, J. and D. Bernheim. 2009. "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica*, **77**(5):1607-1636.

Andreoni, J. and J. Miller. 2003. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*, **70**(2): 737-53.

Benabou, R. and J. Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review*, **96**(5): 1652-1678.

Bettenhausen, K. and J. Murnighan. 1991. "The Development of an Intragroup Norm and the Effects of Interpersonal and Structural Challenges." *Administrative Science Quarterly*, **36**: 20-35.

Bicchieri, C. 2006. The Grammar of Society: the Nature and Dynamics of Social Norms. Cambridge University Press.

Bicchieri, C. and E. Xiao. 2009. "Do The Right Thing: But Only If Others Do So." *Journal of Behavioral Decision Making*, **22**(2): 191-208.

Burks, S. and E. Krupka. (unpublished manuscript). “Behavioral Economic Field Experiments Can Identify Normative Alignments and Misalignments within a Corporate Hierarchy: Evidence from the Financial Services Industry”.

Camerer, C. and E. Fehr. 2004. "Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists." Foundations of Human Sociality -- Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies. Ed. J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr and H. Gintis.

Campo, S., D. Brossard, M.S. Frazer, T. Marchell, D. Lewis and J. Talbot, 2003, “Are Social Norms Campaigns Really Magic Bullets? Assessing the Effects of Students’ Misperceptions of Drinking Behavior.” *Health Communication*, **15**(4): 481-497.

Cialdini, R., R. Reno and C. Kallgren. 1990. "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places." Journal of Personality and Social Psychology, **58**(6): 1015-26.

Charness, G. and M. Dufwenberg. 2006. “Promises and Partnerships.” *Econometrica*, **74** (6): 1579-1601.

Conroy, S. and T. Emerson. 2006. "Changing Ethical Attitudes: The Case of the Enron and ImClone Scandals." *Social Science Quarterly*, **87**(2):395–410.

Dawes, R., J. McTavish and H. Shaklee.1977. “Behavior, communication, and assumptions about other people’s behavior in a commons dilemma situation.” *Journal of Personality and Social Psychology*, **35**: 1-11.

Deutsch, M. and H. Gerard. 1955. "A Study of Normative And Informational Social Influences Upon Individual Judgment." *The Journal of Abnormal and Social Psychology*, **51**(3): 629-36.

Dufwenberg, M. and G. Kirchsteiger. 2000. “Reciprocity and Wage Undercutting.” *European Economic Review*, **44**(4-6):1069-1078.

Dur, R., A. Non and H. Roelfsema. 2010. "Reciprocity and Incentive Pay in the Workplace." *Journal of Economic Psychology*, **31**(4):676-686.

Elster, J. 1989. The Cement of Society: a Study of Social Order. Studies in Rationality and Social Change, Cambridge University Press.

Englmeier, F. and S. Leider. (unpublished manuscript). "Contractual and Organizational Structure with Reciprocal Agents."

Ergeneli, A. 2005. "A Cross-Cultural Comparison of Ethical Behavior in Business Related Dilemmas: A Comparison among Turkish, Egyptian, Kirghiz and Kazak Marketing Employees." *Problems and Perspectives in Management*, **2**:135-147

Falk, A. and M. Kosfeld. 2006. "Distrust - The Hidden Cost of Control". *The American Economic Review*, **96**(5): 1611-1630.

Fehr, E. and S. Gächter. 2000. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives*, **14**:159-81.

Fehr, E., O. Hart and C. Zehnder. 2009a. "Contracts, Reference Points, and Competition - Behavioral Consequences of the Fundamental Transformation." *Journal of the European Economic Association*, **7**:561-572.

Fehr, E., O. Hart and C. Zehnder. 2009b. "Contracts as Reference Points-Experimental Evidence." forthcoming *American Economic Review*.

Fehr E., E. Kirchler, A. Weichbold and S. Gächter. 1998. “When Social Norms Overpower Competition: Gift Exchange in Experimental Labor Markets.” *Journal of Labor Economics*, **16**(2): 324-351

Fehr, E. and A. Falk. 1999. "Wage Rigidity in a Competitive Incomplete Contract Market." *The Journal of Political Economy*, **107**(1): 106-134.

Fehr, E., A. Klein and K. Schmidt. 2007. “Fairness and Contract Design.” *Econometrica*, **75**(1): 121-154.

Fehr, E., S. Kremhelmer and K. Schmidt. 2008. “Fairness and the Optimal Allocation of Ownership Rights”. *Economic Journal*, **118**(531): 1262-1284.

Fehr, E. and K. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *The Quarterly Journal of Economics*, **114**(3): 817-68.

Fehr, E. and K. Schmidt. 2004. “Fairness and incentives in a multi-task principal-agent model.” *Scandinavian Journal of Economics*, **106**(3): 453-474.

Fisher, P. and S. Huddart. 2008."Optimal Contracting With Endogenous Social Norms." *American Economic Review*, **98**(4): 1459-75.

Gächter, S., D. Nosenzo and M. Sefton. (unpublished manuscript). "Peer Effects in Pro-Social Behavior: Social Norms or Social Preferences?"

Gino, F., D.A. Moore and M.H. Bazerman, 2008. "No Harm, No Foul: The Outcome Bias in Ethical Judgement." Harvard Business School Working Paper, No. 08-080.

Hart, O. and J. Moore. 2008. "Contracts as Reference Points." *Quarterly Journal of Economics,* **123** (1): 1-48.

Kandori, M. 1992. "Social Norms and Community Enforcement." *Review of Economic Studies*, **59**: 62-80.

Kessler J. and S. Leider, 2010 "Norm and Contracting." *Management Science*, forthcoming.

Krupka, E., R.Weber and R. Croson. (unpublished manuscript). "When in Rome: Identifying Social Norms as a Group Phenomenon."

Krupka, E. and R.Weber. 2009. "The Focusing and Informational Effects of Norms on Pro-Social Behavior." *Journal of Economic Psychology*, **30**: 307-20.

Lambert, N. and Y. Shoham. 2009. "Eliciting Truthful Answers to Multiple-choice Questions." *EC '09 Proceedings of the tenth ACM conference on Electronic Commerce.*

Leider, S., M. Möbius, T. Rosenblat and Q. Do. 2010. "Directed Altruism and Enforced Reciprocity in Social Networks." *Quarterly Journal of Economics*, **124**(4).

Malhotra, D. and J. Murnighan. 2002. "The Effects of Contracts on Interpersonal Trust." *Administrative Science Quarterly*, **47**.

Masclet, D., C. Noussair, S. Tucker, M-C. Villeval. 2003. "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism," *American Economic Review*, **93**(1):366-380.

McKinney, J., and C. Moore. 2008. "International Bribery: Does a Written Code of Ethics Make a Difference in Perceptions of Business Professionals." *Journal of Business Ethics*, **79**:103-111.

Mehta, J., C. Starmer and R. Sugden. 1994. "The Nature of Salience: An Experimental Investigation of Pure Coordination Games." *American Economic Review*, **84**(3): 658-73.

Ostrom, E. 2000. "Collective Action and the Evolution of Social Norms." *Journal of Economic Perspectives*, **14**(3):137-58.

Oumlil A. and J. Balloun, 2009. "Ethical Decision-Making Differences Between American and Moroccan Managers." *Journal of Business Ethics*, **84**:457–478

Perkins, H. and H. Wechsler. 1996. "Variation in Perceived College Drinking Norms and its Impact on Alcohol Abuse: A Nationwide Study." *Journal of Drug Issues*, **26**(4):961-974.

Rigdon, M. 2009. “Trust and Reciprocity in Incentive Contracting.” *Journal of Economic Behavior and Organization,* **70**: 93-105.

Sally, D.1992. “Conversation and cooperation in social dilemmas: Experimental evidence from 1958 to 1992.” *Rationality and Society*, 7(1):58-92.

Scott, R. 2003. “A Theory of Self-Enforcing Indefinite Agreements.” *Columbia Law Review,* **103**(7): 1641-1699.

Schelling, T. 1960. The Strategy of Conflict. Cambridge, MA, Harvard University Press.

Schwartz, S. 1973. “Normative Explanations for Helping Behavior: A Critique, Proposal and Empirical Test.” *Journal of Experimental Social Psychology*, **9**(4): 349-364.

Seinena, I. and A. Schramb. 2006. “Social status and group norms: Indirect reciprocity in a repeated helping experiment.” *European Economic Review*, **50**(3):581-602.

Sliwka, D. 2007. “Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes.” *American Economic Review*, 97 (3): 999-1012.

Sugden, R. 1995. "A Theory of Focal Points." *The Economic Journal*, **105**(430):533-50.

Jean Tirole. 1999. “Incomplete Contracts: Where Do We Stand?”, *Econometrica*, **67**(4):741-781.

Vuong, Q. 1989. “Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses.” *Econometrica*, **57**(2): 307-333.

Young, P. 1998. “Social Norms and Economic Welfare.” *European Economic Review,* **42**: 821-30.

Figure 1: Screen shot of the decision screen subjects saw in Module 1 (injunctive norm elicitation).

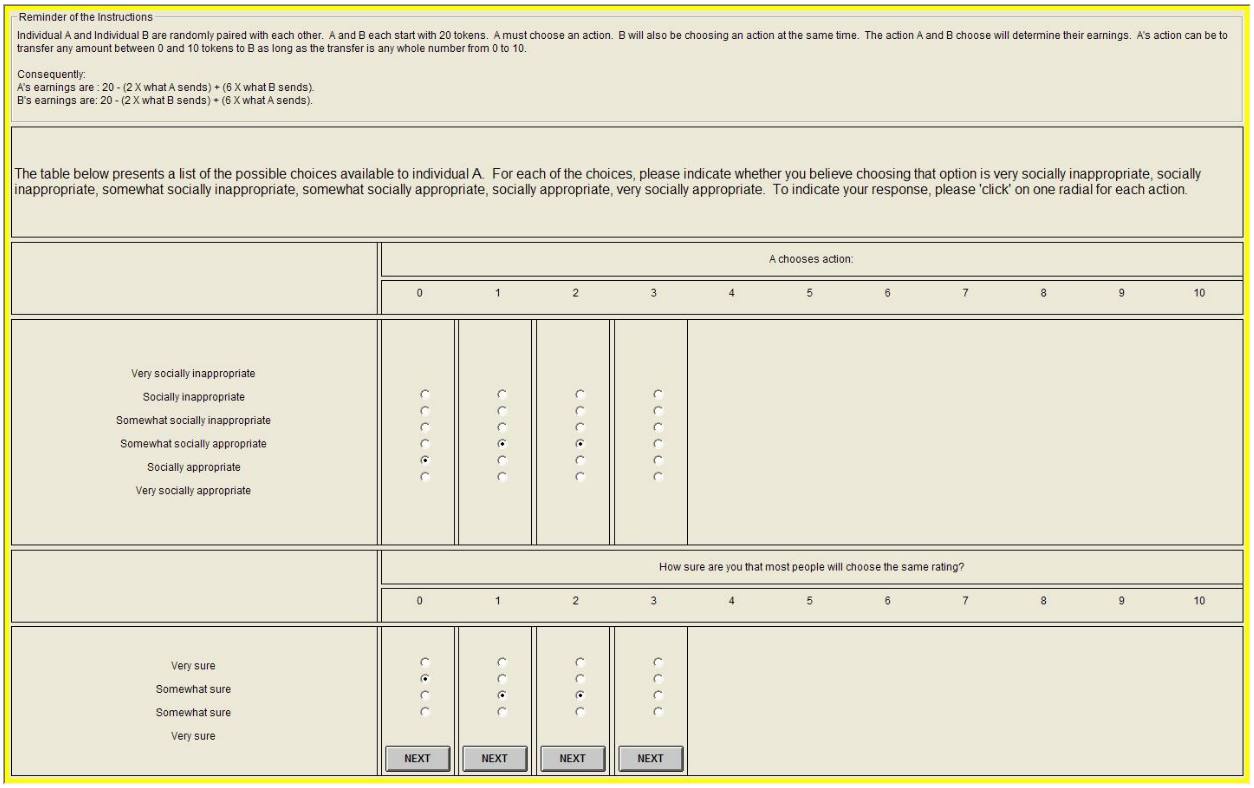


Figure 2: Average appropriateness ratings for the Double Dictator Game with and without agreement.

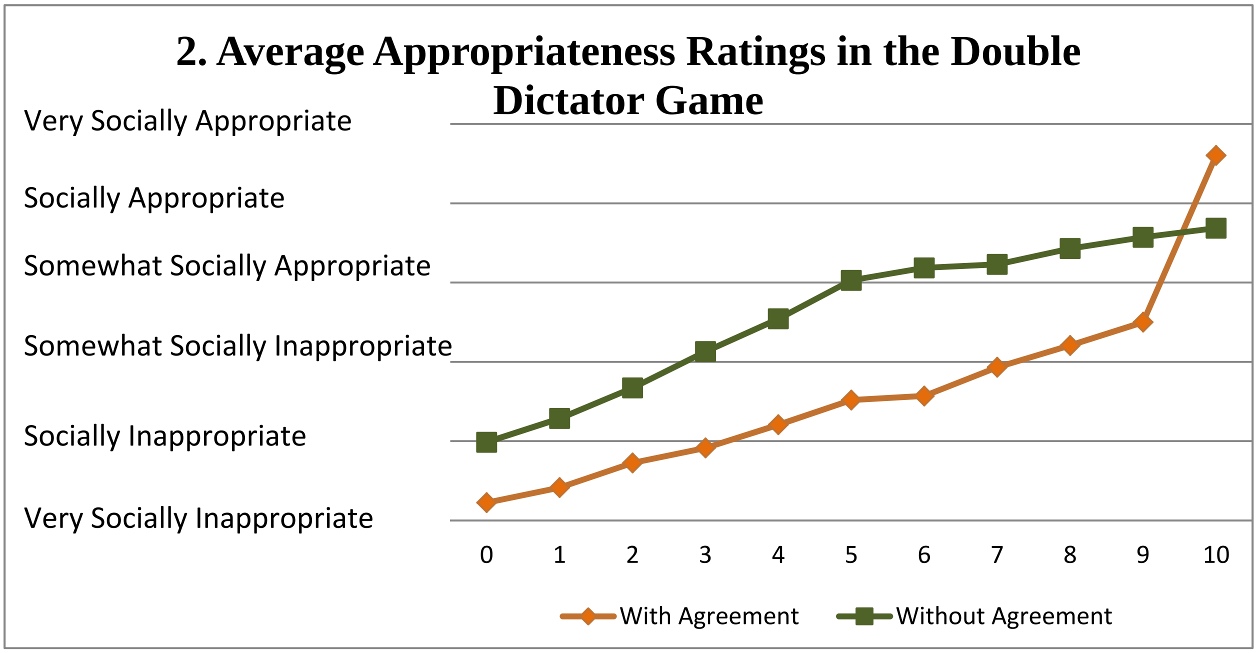


Figure 3: Average appropriateness ratings for the Bertrand Game with and without agreement.

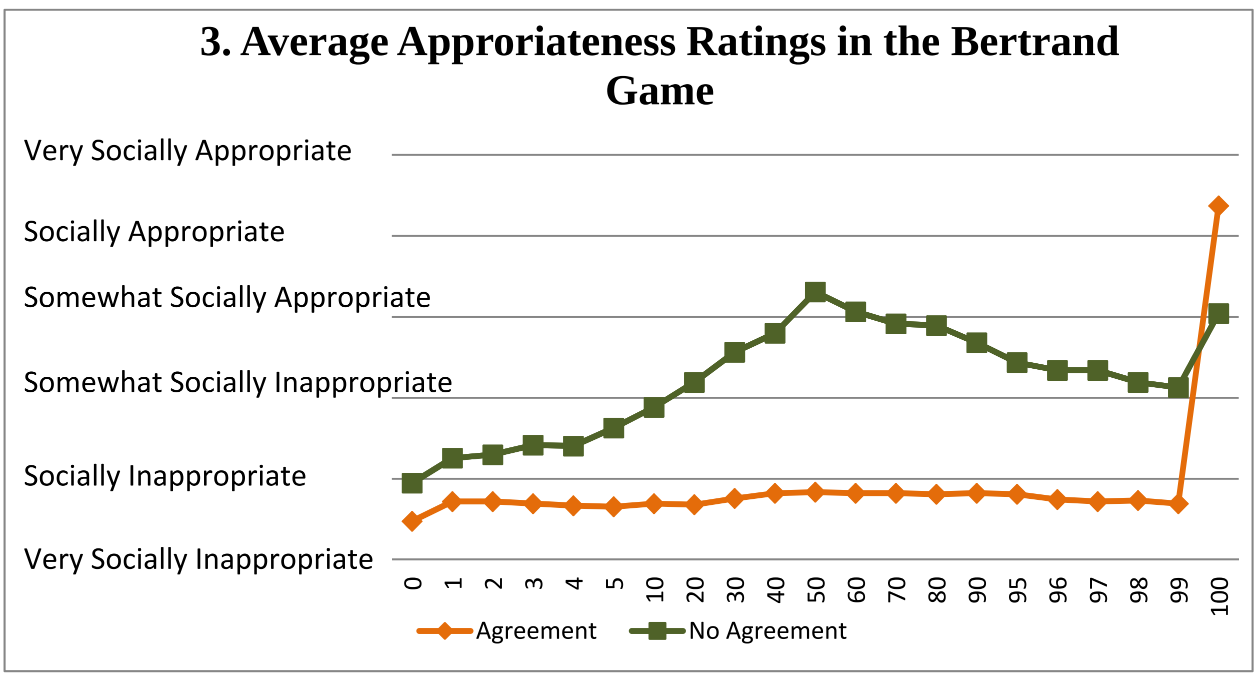
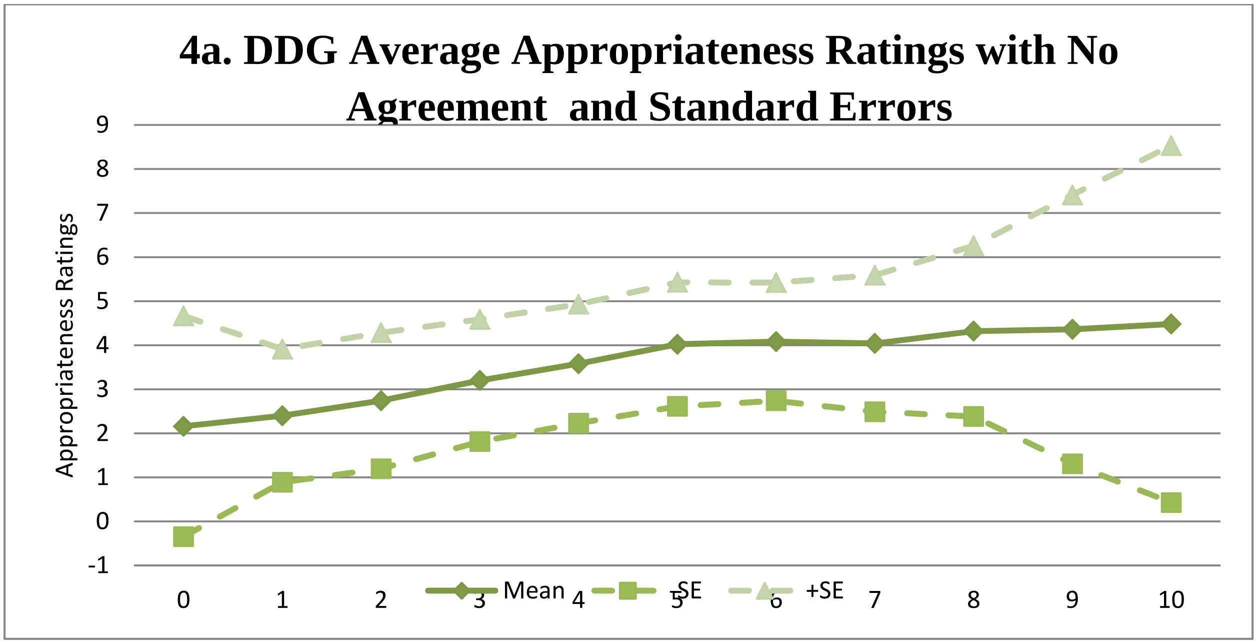


Figure 4a-4b: Standard Errors around the average appropriateness ratings for the Double Dictator Game with and without agreement.



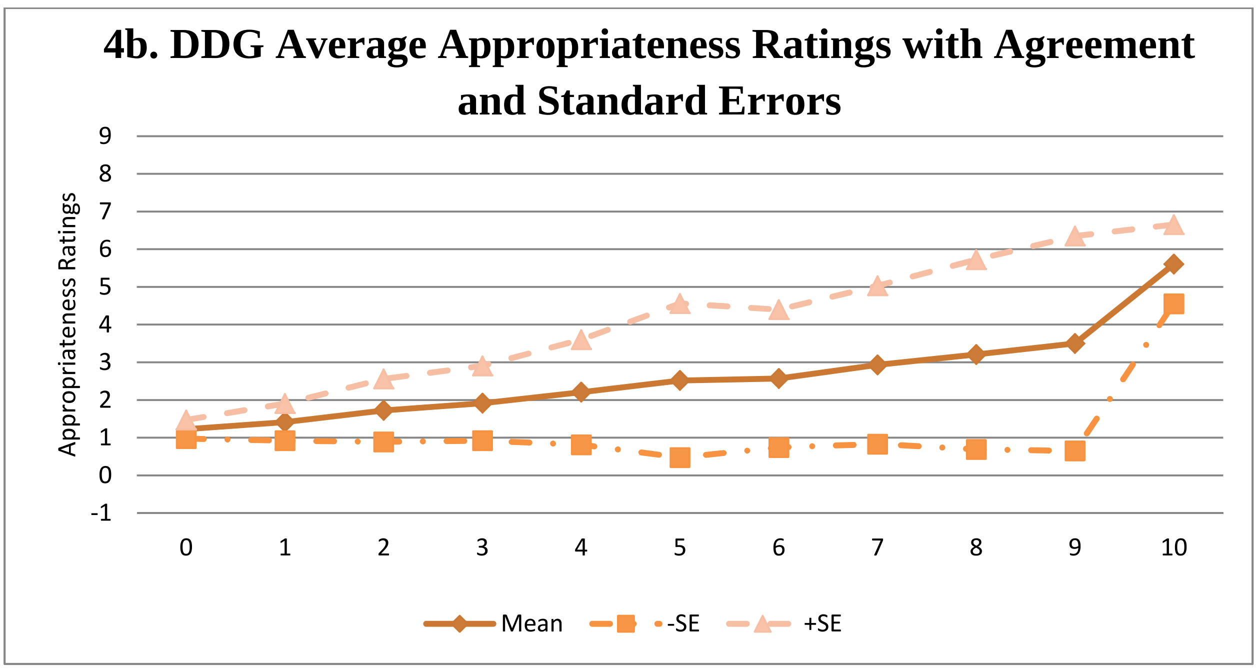
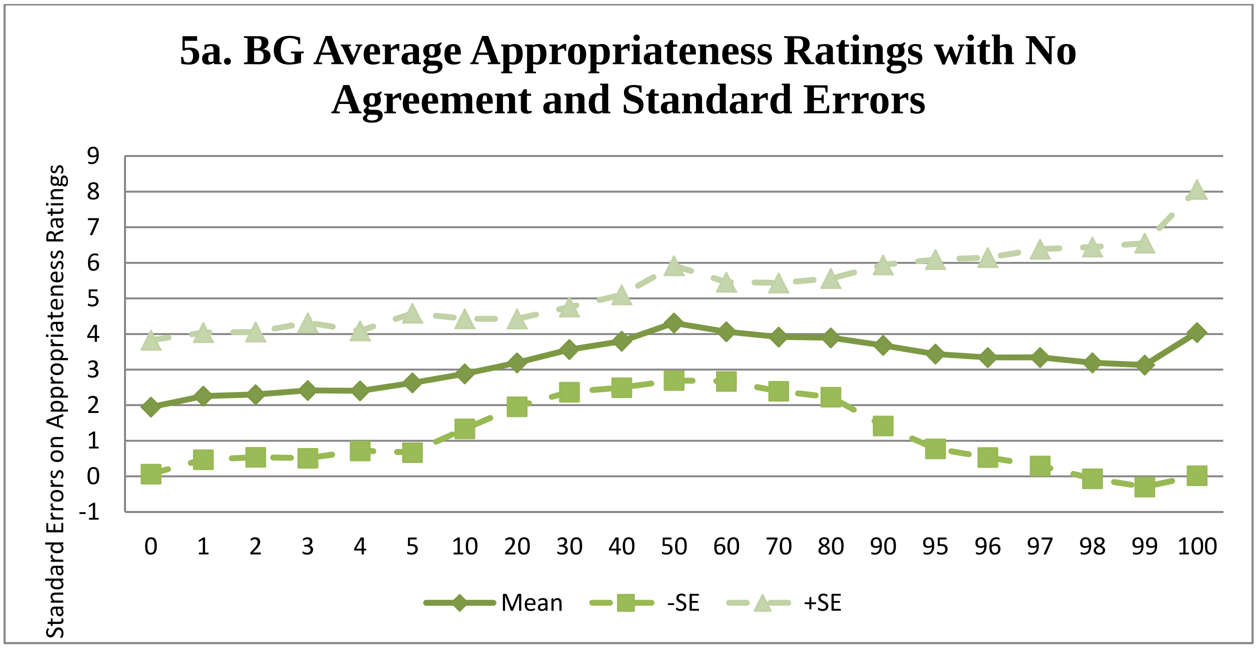


Figure 5a-5b: Standard Errors around the average appropriateness ratings for the Bertrand Game with and without agreement.



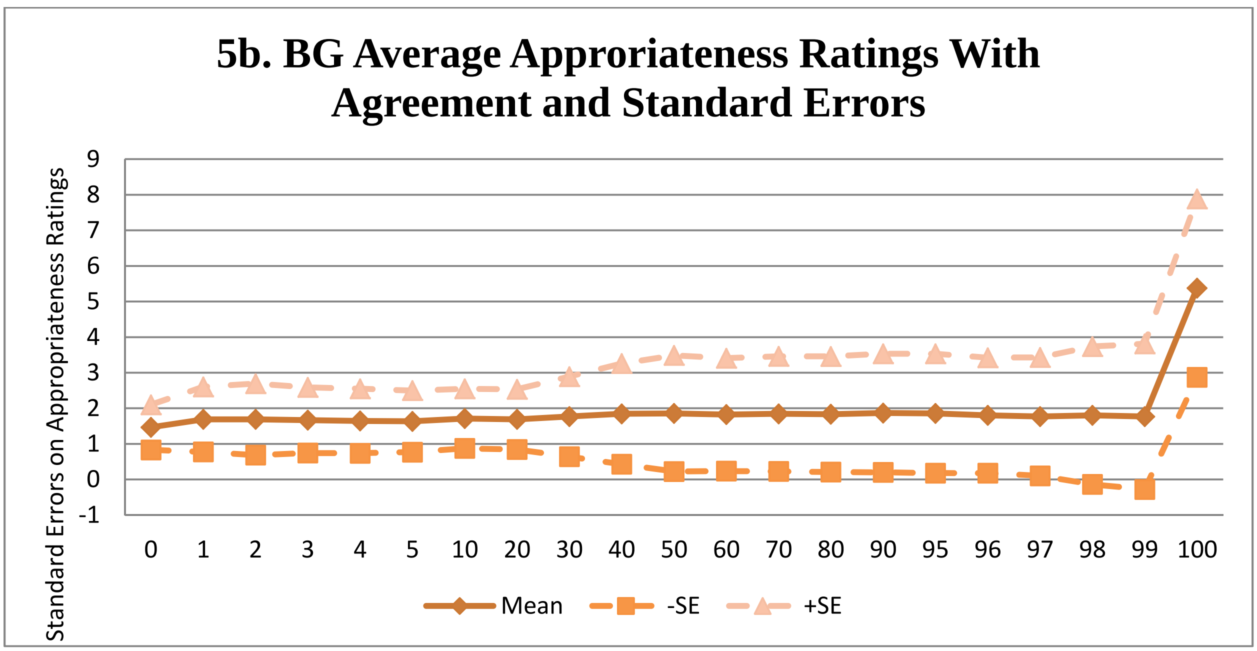


Figure 6: Actual and Predicted Choices for the Double Dictator Game

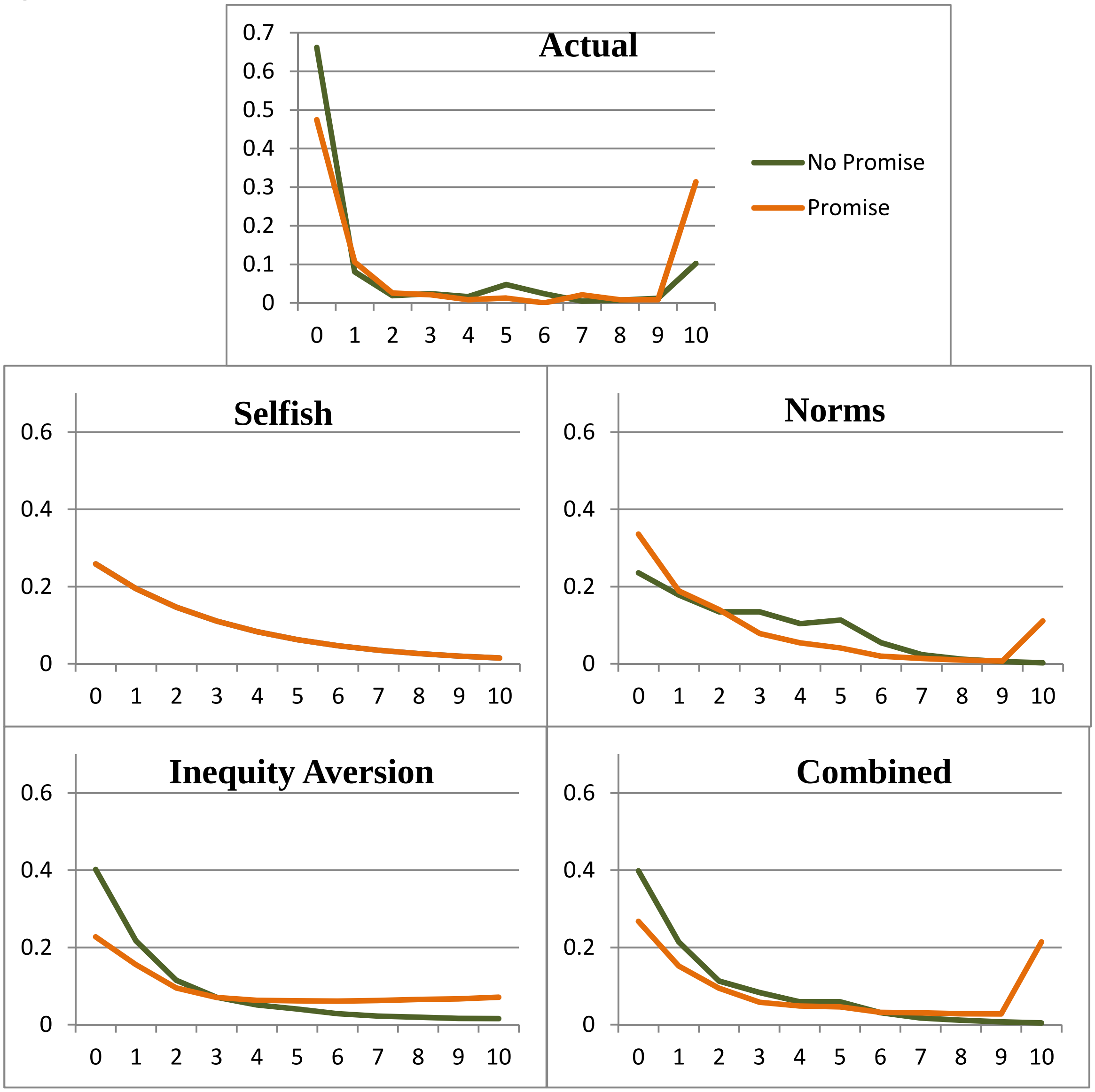


Figure 7: Actual and Predicted Choices for the Bertrand Game

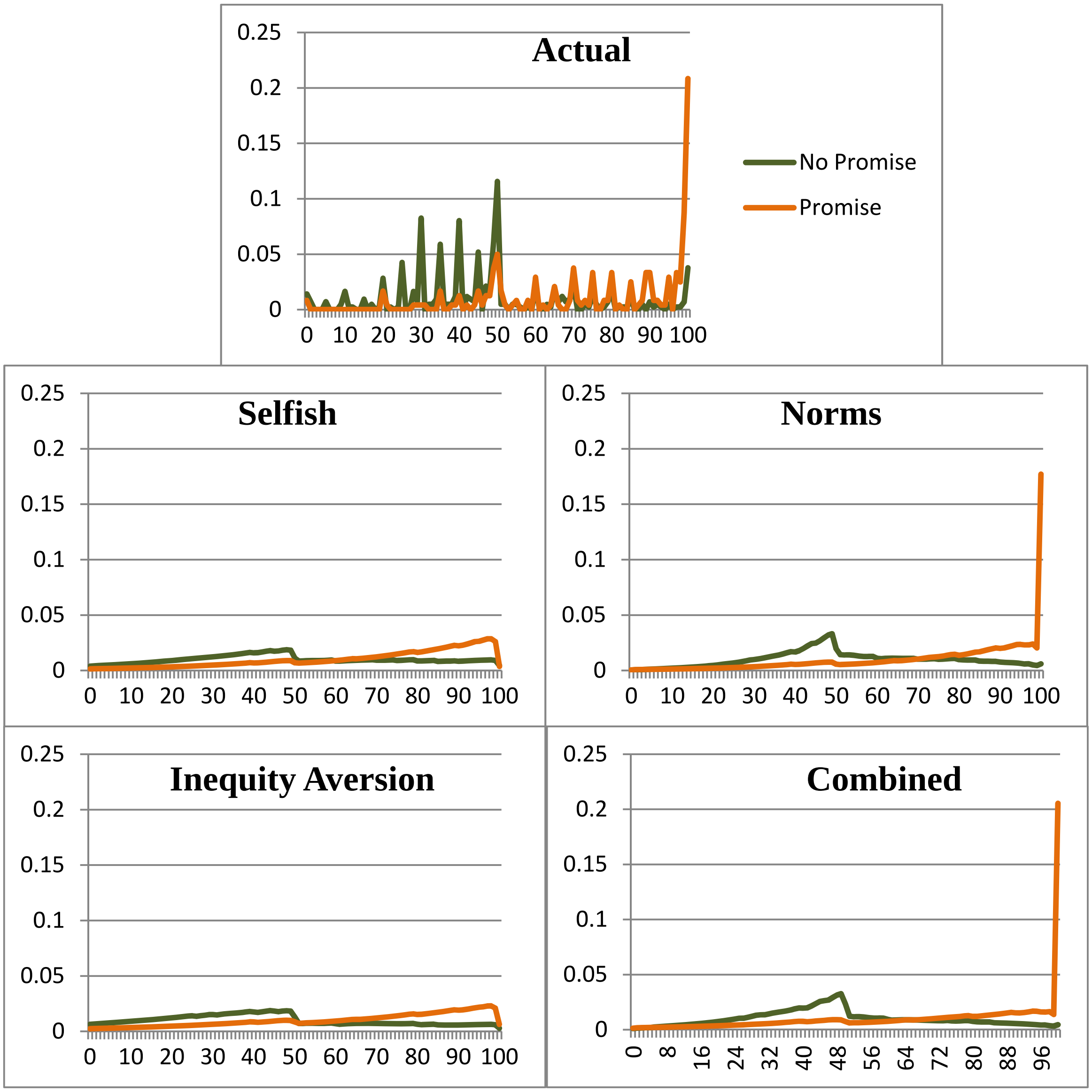


Table 1: Experiment Overview

	Double Dictator Game		Bertrand Game	
	With Agreement	Without Agreement	With Agreement	Without Agreement
# Subjects	84	90	90	94
# Sessions	9	9	9	9
Average Payoff	$28.17	$28.37	$32.61	$29.65

Table 2: OLS regressions on appropriateness ratings for the Injunctive Norm Elicitation for the Double Dictator game and the Bertrand Game.

	DDG			BG		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Action	0.275***	0.302***	0.300***	0.0758***	0.0749***	0.0740***
	(0.0147)	(0.0155)	(0.0155)	(0.00540)	(0.00558)	(0.00552)
Agreement	-1.380***	-0.962***	-0.937***	-1.106***	-0.780***	-0.736***
	(0.143)	(0.151)	(0.151)	(0.116)	(0.126)	(0.126)
Agreement × Action	0.0781***	-0.0263	-0.0241	-0.0178**	-0.0623***	-0.0615***
	(0.0187)	(0.0190)	(0.0191)	(0.00781)	(0.00660)	(0.00655)
Highest Action		-0.590***	-0.529**		0.0726	0.139
		(0.207)	(0.208)		(0.201)	(0.201)
Agreement × Highest Action		2.298***	2.248***		3.423***	3.362***
		(0.254)	(0.255)		(0.274)	(0.274)
Middle Action			0.457***			1.145***
			(0.109)			(0.121)
Agreement × Middle Action			-0.373**			-1.041***
			(0.153)			(0.150)
Constant	2.017***	1.910***	1.879***	2.391***	2.398***	2.350***
	(0.122)	(0.131)	(0.132)	(0.0988)	(0.108)	(0.107)
Observations	1914	1914	1914	3864	3864	3864
# of Subjects	174	174	174	184	184	184

Notes: Dependent Variable is the norm rating for each action; Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table 3: OLS regression testing for changes in the shape of the social norm profile.

	Top Jump		Middle Jump	
	DDG	BG	DDG	BG
VARIABLES	(1)	(2)	(3)	(4)
Agreement	1.883***	2.696***	-0.0921	-0.355***
	(0.207)	(0.268)	(0.0831)	(0.0904)
Constant	0.0333	0.915***	0.211***	0.378***
	(0.0843)	(0.175)	(0.0707)	(0.0851)
Observations	174	184	174	184
R-squared	0.335	0.358	0.335	0.076

Notes: Dependent Variable for (1) and (2) is the difference in norm rating for the highest and second highest action, for (3) and (4) it is the difference between the norm rating for the middle action and the average rating of the actions one higher and one lower; Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Conditional Logit regression testing for changes in the shape of the social norm profile.

	DDG				BG			
VARIABLES	(1)	(2)	(3)	(4)	(4)	(5)	(5)	(6)
Action Payoff	0.142***	0.484***	0.0775***	0.322***	0.0470***	0.0516***	0.0514***	0.0515***
	(0.00766)	(0.0343)	(0.00982)	(0.0300)	(0.00225)	(0.00243)	(0.00259)	(0.00552)
		[0.131]		[0.0722]		[.000824]		[.000828]
Norm Rating		1.934***		1.300***		1.204***		1.099***
		(0.171)		(0.142)		(0.0504)		(0.0497)
		[0.723]		[0.377]		[.0632]		[.0587]
Inequity Aversion			-0.0458***	-0.0440***			-0.0201***	-0.0188***
			(0.00249)	(0.0025)			(0.0026)	(0.00214)
				[0.00274]				[.00159]
Observations	656	656	656	656	664	664	664	664
Log Likelihood	-1360.1	-1297.4	-1089.2	-1044.8	-2786.1	-2617.4	-2741.9	-2567.2
Bayesian IC	2729.0	2598.8	2196.2	2116.3	5583.2	5257.0	5506.2	5167.7

Notes: The Dependent Variable is the chosen action in the Double Dictator Game or Bertrand Game; Standard errors are reported in parentheses, with bootstrapped standard errors in brackets for specifications with norm ratings; *** p<0.01, ** p<0.05, * p<0.1.