For online publication

Appendix I: Supporting Materials

A Meeting of the Minds:
Informal Agreements and Social Norms

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)

November 4, 2015

Contents

1 Experiment Design Details 2
1.1 Design overview 2
1.2 Screen shot of the norm elicitation screen 3
1.3 Payments and participation rate 3
2 Additional Analysis of Norm Ratings 4
2.1 The distribution of injunctive norm ratings from Module 1 4
3 Robustness Checks 7
3.1 Summary of robustness checks 7
4 Additional Analysis 13
4.1 Comparing Lying Aversion Specifications 13
4.2 Combining Lying Aversion with Inequality Aversion 13

[1] Corresponding author: ekrupka@umich.edu

1 Experiment Design Details

1.1 Design overview

Below is a table that gives an overview of all experimental treatments and the modules which our subjects participated in.

Table S1: Overview of the Experimental Design

<table>
<tr><td></td><td></td><td colspan="5">← With-in subject →</td></tr>
<tr><td></td><td>Modules / Agreement Environment</td><td>Module 1 Injunctive Norm I</td><td>Module 2 Descriptive Norm</td><td>Module 3 Observation</td><td>Module 4 Injunctive Norm II</td><td>Module 5 Control variables</td></tr>
<tr><td rowspan="2">↕ Between-subject</td><td rowspan="3">No Agreement</td><td rowspan="3">Coordinate appropriateness ratings

$ Incentives</td><td>Version A: Guess mode, max.and min. transfer

$ Incentives</td><td>Either:
Random whole distribution
Random w. over sample of upper tail
Random w. over sample of lower tail</td><td rowspan="3">Coordinate appropriateness ratings

$ Incentives</td><td rowspan="3">All:

1) Altruism measure with ‘helping game’

2) “Lying game”

3) Demographics</td></tr>
<tr><td>Version B: Guess average FOB and SOB

$ Incentives</td></tr>
<tr><td rowspan="2">Informal agreement</td><td rowspan="2">Coordinate appropriateness ratings

$ Incentives</td><td>Version A: Guess mode, max.and min. transfer

$ Incentives</td><td>Either:
Random whole distribution
Random w. over sample of upper tail
Random w. over sample of lower tail</td><td rowspan="2">Coordinate appropriateness ratings

$ Incentives</td><td rowspan="2">All:

1) Altruism measure with ‘helping game’

2) “Lying game”

3) Demographics</td></tr>
<tr><td>Version B: Guess average FOB and SOB

$ Incentives</td></tr>
</table>

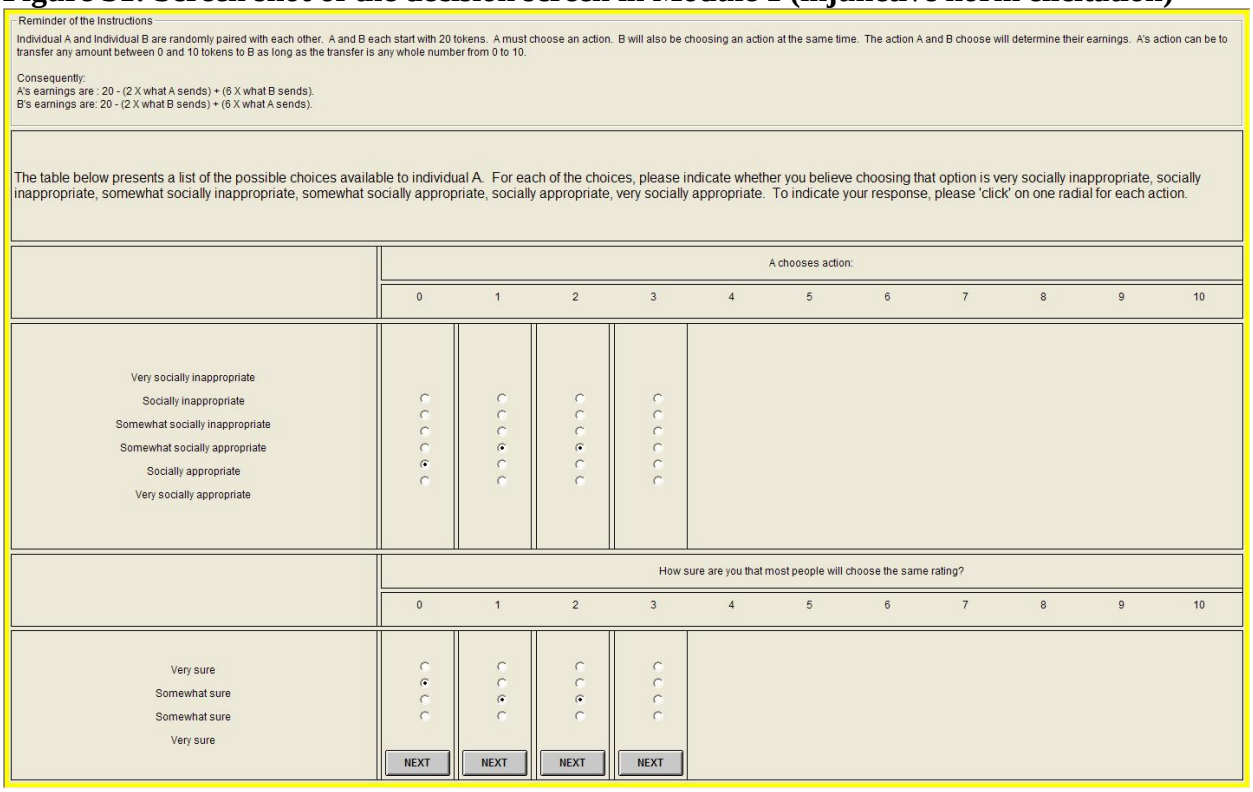
The experimental design consists of five modules. The first module, “Injunctive Norms I”, elicits subjects’ beliefs about normative expectations, and in aggregate, identifies the social norm for that decision context. In the second module, “Descriptive Norms A”, subjects are asked to guess the modal transfer amount and the percent of subjects who transferred the minimum and maximum amount. “Descriptive Norms B”subjects are asked to guess the average first order and second order beliefs of subjects in the KL experiments. Thus, the second module elicits important moments of a subject’s belief about the descriptive norm. In the third module, “Observation”, subjects see the token transfer decisions of five other subjects who participated in the Kessler and Leider study (observations are chosen randomly from the whole distribution of actions, randomly

from the whole distribution with upper-quartile bias, or randomly from the whole distribution with lower-quartile bias). Thus, the third module gives subjects information about the actual distribution of the descriptive norm and, depending on where in the distribution the observations are drawn, is in accordance or at odds with a subject's beliefs about that distribution. The fourth, "Injunctive Norms II", module elicits the injunctive norm again for the same decision context. The fifth module elicits variables we wish to control and includes incentivized measures of altruism and demographic information. Regardless of the treatment condition, all subjects participate in all modules 1-5 and the order in which subjects see modules 1 through 5 is always the same. In all cases, subjects are informed of their individual earnings only after all experimental modules have been completed.

1.2 Screen shot of the norm elicitation screen

Figure S1 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 S1: Screen shot of the decision screen in Module 1 (injunctive norm elicitation)



1.3 Payments and participation rate

A total of 358 participants (in 36 sessions) were recruited in the norm elicitation experiment. A total of 62 participants (in 4 sessions) were recruited in the behavior experiments. 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 S2 details participation and average payoffs by treatment.

Table S2: Experiment participation per session and payouts

	Norm Elicitation Experiments			
	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
	DDG and BG Choice and Second Order Belief Elicitation			
# Subjects	62			
# Sessions	4			
Average Payoff	$16.63			

2 Additional Analysis of Norm Ratings

2.1 The distribution of injunctive norm ratings from Module 1

The presence of an agreement also affects the variance in 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 which may plausibly affect how much concern, γ from equation (1), an individual places on complying with social norms relative to monetary payoffs. The Figures S2-S5 depict the standard errors for each game and treatment. In both games we find that the variance in ratings increases 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 (i.e., what the utility payoff will be from taking these actions). In particular, the rating for 'action 10' in the Double Dictator Game has a higher variance in the No Agreement treatment than each of the ratings for 'action 8' or lower actions (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).

However, in the Double Dictator Game treatment with Agreement we find much more variance in the ratings about the appropriateness of intermediate actions, but less variance about

the appropriateness of 'action 10'. We also see that there is significantly less disagreement about the appropriateness of 'action 10' 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.

Figure S2: Standard errors around the average appropriateness ratings for the Double Dictator Game with no agreement (data from Module 1)

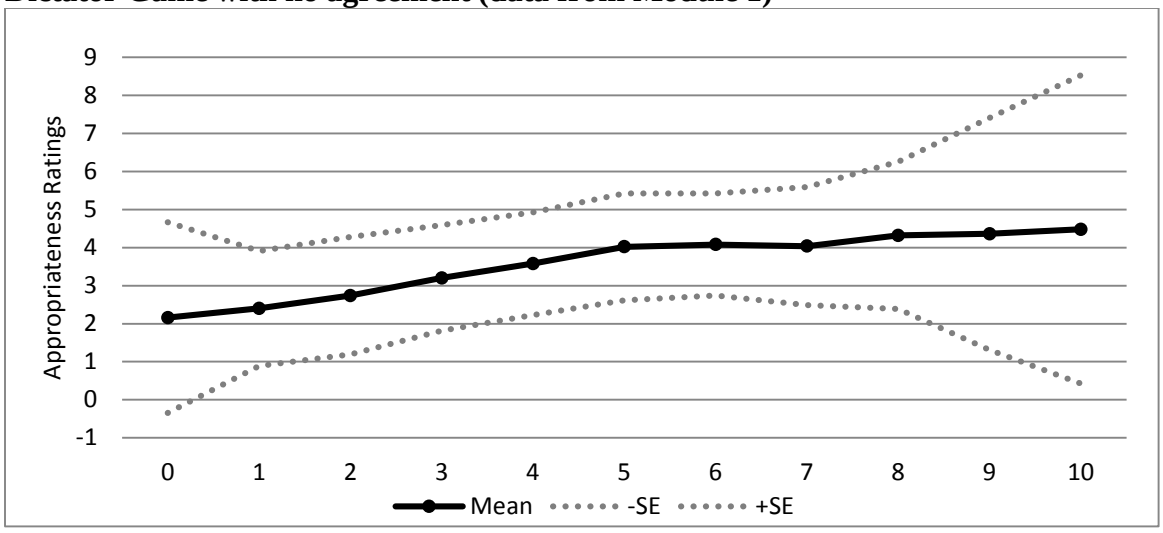


Figure S3: Standard errors around the average appropriateness ratings for the Double Dictator Game with agreement (data from Module 1)

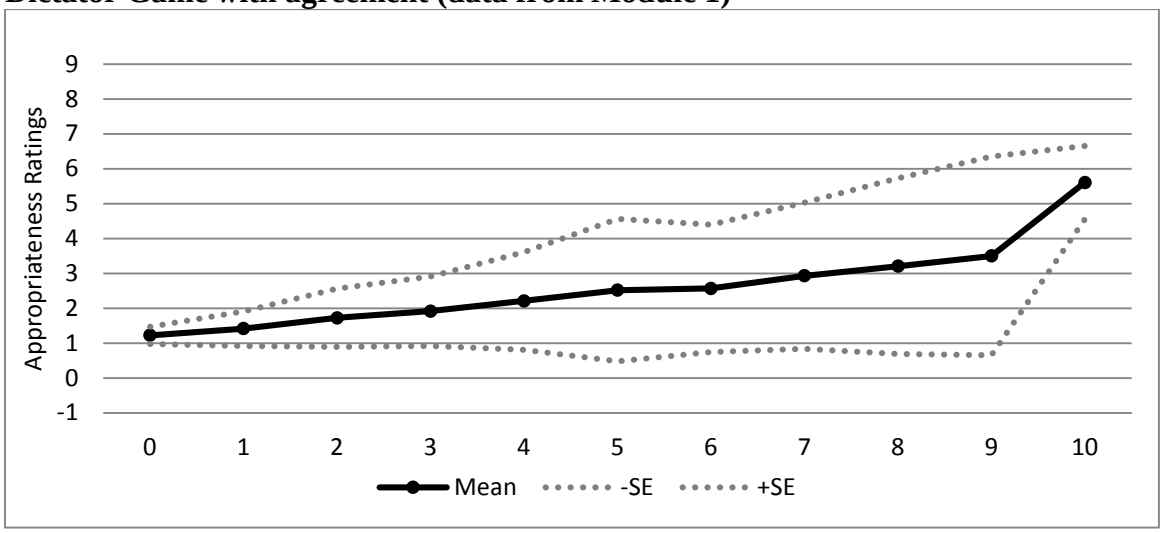


Figure S4: Standard errors around the average appropriateness ratings for the Bertrand Game with no agreement (data from Module 1)

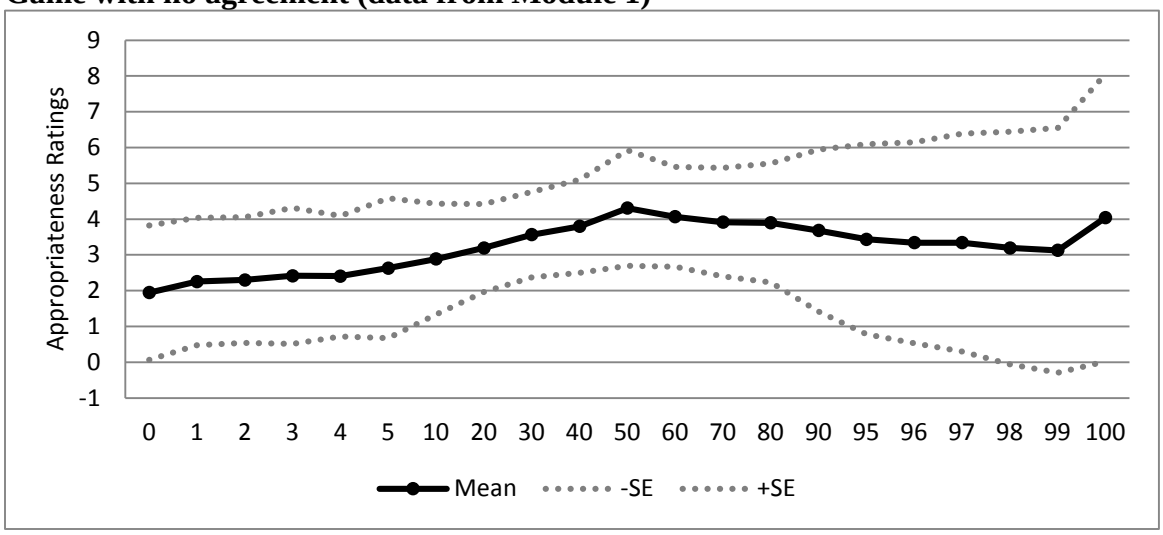
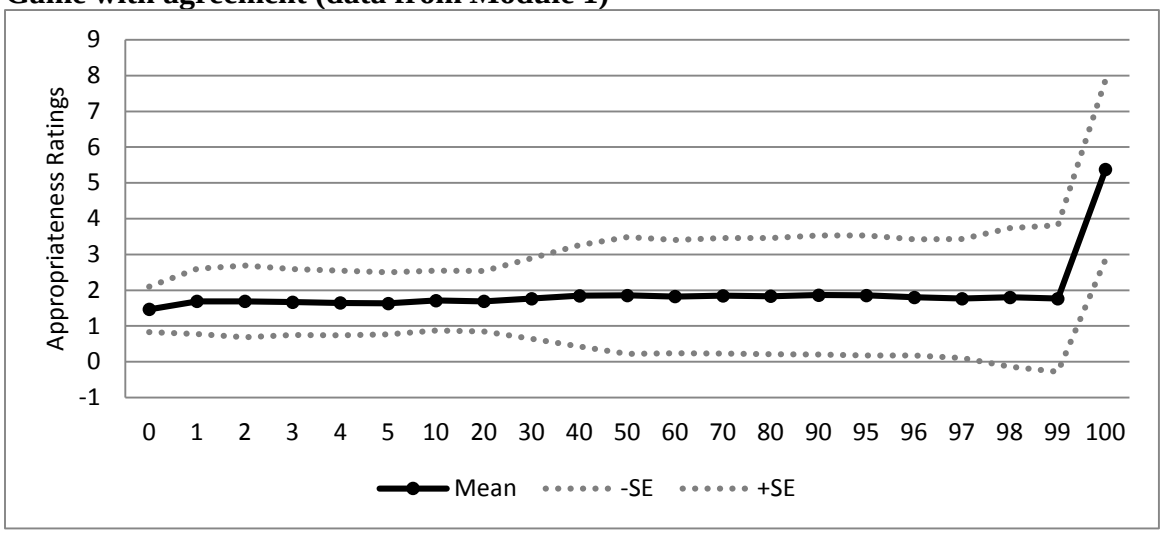


Figure S5: Standard errors around the average appropriateness ratings for the Bertrand Game with agreement (data from Module 1)



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. 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 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 Bertrand Game this is 21

percent and 83 percent respectively (test of proportions: $p < 0.01$ for both). These findings clearly demonstrate that informal 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.

We analyze the effect of the Agreement treatment on the shape of the appropriateness profile more formally in Table S3. Here we construct two summary measures for the jump in ratings for the highest action, and for the middle action. 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 S3 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.

Table S3: 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.

3 Robustness Checks

3.1 Summary of robustness checks

In order for the coordination task to accurately measure the effect of the agreement on the norms profile, and thus have the measured norm profile be useful for predicting behavior, we would like the responses to the coordination task to be robust to competing focal points: Ratings ought not be overly sensitive to individual characteristics or beliefs about the descriptive norm

(what people actually do) and they ought to be sensitive to whether an agreement is in place and not whether the agreement is Pareto efficient.

If we divide subjects into two categories, for each game and each treatment condition, based on their stated belief (elicited in Module 2) about the modal action taken in the original KL experiment we can calculate the average appropriateness ratings (from Module 1) for each action for those two groups.

Table S4: OLS regressions on appropriateness ratings with beliefs for the Injunctive Norm Elicitation for the Double Dictator game and the Bertrand game.

	DDG		BG	
VARIABLES	(1)	(2)	(3)	(4)
Action	0.300***	0.300***	0.0740***	0.0740***
	(0.0155)	(0.0155)	(0.00552)	(0.00552)
Agreement Treatment	-0.937***	-0.805***	-0.736***	-0.540***
	(0.151)	(0.157)	(0.126)	(0.165)
Agreement × Action	-0.0241	-0.0241	-0.0615***	-0.0615***
	(0.0191)	(0.0191)	(0.00655)	(0.00656)
Highest Action	-0.529**	-0.529**	0.139	0.139
	(0.208)	(0.210)	(0.201)	(0.203)
Agreement × Highest Action	2.248***	2.248***	3.362***	3.362***
	(0.255)	(0.258)	(0.274)	(0.277)
Middle Action	0.457***	0.457***	1.145***	1.145***
	(0.109)	(0.110)	(0.121)	(0.121)
Agreement × Middle Action	-0.373**	-0.373**	-1.041***	-1.041***
	(0.153)	(0.153)	(0.150)	(0.149)
Beliefs		-0.0971**		-0.121**
		(0.0481)		(0.0616)
Constant	1.879***	2.205***	2.350***	2.830***
	(0.132)	(0.228)	(0.107)	(0.276)
Observations	1914	1914	3864	3864
Number of subject	174	174	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 S4 reports tis OLS regression. We find that subjects with more optimistic beliefs (high mode) about what others will do tend to have lower ratings for low actions, and higher ratings for high actions. However, the shapes of the norm functions are preserved and major changes in the shape of the norm function in the presence of agreements are not driven by beliefs.

Observing others' behavior can create focal points as well. Module 3 shows our subjects five randomly drawn actions from what other subjects actually did in KL's experiment. After

observing what others actually did, our subjects perform the same injunctive norm rating tasks as in Module 1. We divided our subjects up into whether they observed draws that were "upper tail biased", "lower tail biased" or "random from the whole distribution" and then test whether (a) observing others' behavior affects a subject's norm ratings and (b) whether elicited beliefs about norms are sensitive to observing behavior that is *inconsistent* with one's beliefs about the descriptive norm.

Table S5: OLS regression testing for changes in the shape of the social norm profile in the second norm elicitation module after different observations

	DDG	BG
VARIABLES	(1)	(2)
Agreement Treatment	-1.101***	-0.532**
	(0.236)	(0.209)
Action	0.238***	0.0835***
	(0.0322)	(0.0156)
Agreement × Action	0.0339	-0.0442**
	(0.0392)	(0.0197)
Highest Action	-0.101	-0.191
	(0.110)	(0.165)
Agreement × Highest Action	1.562***	3.083***
	(0.219)	(0.287)
Middle Action	0.443***	1.453***
	(0.0790)	(0.117)
Agreement × Middle Action	-0.494***	-1.234***
	(0.100)	(0.150)
High Observation	-0.153	0.163
	(0.137)	(0.140)
Low Observation	0.0737	0.0909
	(0.149)	(0.116)
Constant	2.400***	2.038***
	(0.205)	(0.181)
Observations	1,914	3,864
Number of subjects	174	184

Notes: Dependent variable is the norm rating for each action in the second norm elicitation module; robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table S5 reports a regression testing the effect of "observation" on responses in the social norms elicitation module of our experiment. The types of observation ('high observation' or 'low observation') are dummy variables where the omitted category is those subjects who observations that were drawn randomly from the whole distribution. The dependent variable is the norms ratings our subjects gave us in Module 4. The results show that different observations have no significant effect on overall appropriateness ratings in the second injunctive norm elicitation stage (p=0.149,

p=0.137 for 'high-' and 'low-observation' dummy variables in the Double Dictator Game and p=0.140, p=0.116 in the Bertrand Game respectively).

We use subject choices in the 'Advice Game'[2] and the 'Helping Game'[3] to create proxy measures of individual norm-compliance to test whether responses in the norm elicitation module differ by whether someone is particularly norm-sensitive. Table S6 summarizes the regression results for the 'Advice Game' (Column (1) and (3)) and 'Helping Game' (Column (2) and (4)). We create dummy variables based on these categorizations and find that there is no significant difference in the injunctive norm appropriateness ratings from Module 1 among the player types in the 'Advice Game' in either the Double Dictator Game ("lair", "altruistic lair", and "strategic honest" respectively, p=0.760, p=0.613, p=0.715) or the Bertrand Game ("lair", "altruistic lair", and "strategic honest" respectively, p=0.802, p=0.944, p=0.707). There is also no significant difference in appropriateness ratings among player types in the 'Helping Game' in the Double Dictator Game ("high" and "middle" helper respectively, p=0.379, p=0.371) or in the Bertrand Game ("high" and "middle" helper respectively p=0.221, p=0.287).

[2] In the 'Advice Game', 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 characterize first movers in the 'Advice Game' who send an honest message and believe that the second mover will follow their advice as "Honest", those who send an honest message but do not believe that the second mover will follow their advice as "strategic honest", those who send a message that is a lie and believe that the second mover will take their advice as "liars", and those who send a lying message but who do not believe that the second mover will take their advice as "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).

[3] In 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. 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". The numbers of subjects of type "honest", "liar", "altruistic lair", and "strategic honest" in Double Dictator Game are 24, 26, 9, and 28 respectively; in Bertrand Game, we have 27, 14, 16, and 35 respectively. The numbers of subjects of type "low helper", "middle helper", and "high helper" in Double Dictator Game are 21, 49 and 17 respectively; in the Bertrand Game we have 24, 51 and 17 respectively). Thus, higher actions of helpers in the 'Helping Game' can be used to characterize subjects who may feel more strongly about adhering to social norms of helping than those helpers who choose lower actions.

Table S6: OLS regression testing for changes in the shape of the social norm profile with different types of players in 'Advice' and 'Helping Game'

	DDG		BG	
VARIABLES	(1)	(2)	(3)	(4)
Agreement Treatment	-0.959***	-0.573*	-0.612**	-0.631**
	(0.359)	(0.338)	(0.276)	(0.291)
Action	0.304***	0.327***	0.0762***	0.0880***
	(0.0475)	(0.0460)	(0.0213)	(0.0230)
Agreement X Action	-0.00236	-0.0286	-0.0598**	-0.0602**
	(0.0558)	(0.0550)	(0.0262)	(0.0271)
Highest Action	-0.503***	-0.353***	-0.0952	-0.273
	(0.161)	(0.132)	(0.269)	(0.216)
Agreement X Highest Action	1.889***	1.576***	3.228***	3.376***
	(0.326)	(0.306)	(0.442)	(0.413)
Middle Action	0.416***	0.394***	1.241***	1.288***
	(0.144)	(0.131)	(0.171)	(0.165)
Agreement X Middle Action	-0.333*	-0.202	-1.100***	-1.129***
	(0.175)	(0.160)	(0.205)	(0.204)
Strategic Honest	-0.0699		0.0526	
	(0.229)		(0.210)	
Strategic Lair	-0.123		-0.0178	
	(-0.243)		(0.255)	
Altruistic Lair	0.137		0.0900	
	(0.375)		(0.240)	
Low helper		0.379		0.221
		(0.293)		(0.272)
Middle helper		0.371*		0.287
		(0.214)		(0.211)
Constant	1.976***	1.387***	2.254***	2.037***
	(0.333)	(0.338)	(0.246)	(0.267)
Observations	957	957	1,932	1,932
Number of Subjects	87	87	92	92

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 S7 reports regression testing the effect of individual characteristics on injunctive norm ratings. There is no significant gender difference in appropriateness ratings (p=0.215 for gender dummy in the Double Dictator Game; p=0.883 in the Bertrand Game).

Table S7: OLS regression testing for gender difference

	DDG	BG
VARIABLES	(1)	(2)
Agreement Treatment	-0.970***	-0.737***
	(0.224)	(0.202)
Action	0.298***	0.0732***
	(0.0331)	(0.0159)
Agreement × Action	-0.0224	-0.0609***
	(0.0393)	(0.0189)
Highest Action	-0.541***	0.114
	(0.109)	(0.190)
Agreement × Highest Action	2.259***	3.432***
	(0.244)	(0.298)
Middle Action	0.465***	1.154***
	(0.101)	(0.126)
Agreement × Middle Action	-0.381***	-1.046***
	(0.123)	(0.144)
Male	0.151	0.0178
	(0.122)	(0.121)
Constant	1.826***	2.350***
	(0.191)	(0.182)
Observations	1,903	3,780
Number of subjects	173	180

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

We also ran two additional norm elicitation sessions for the Double Dictator Game and the Bertrand Game where subjects agree on an action *other* than the Pareto efficient one. This tests the effect of an informal agreement on a non-Pareto efficient action on the social norms profile. In the Double Dictator Game we elicited social norms profiles from subjects who read about a scenario in which the agreement between the two parties was to transfer 7 tokens and in the Bertrand Game they read a scenario in which the individuals agreed to take action 70.

We find that agreeing on ‘action 7’ / ‘action 70’ makes it significantly more appropriate to take ‘action 7’ / ‘action 70’ than when there is no agreement (p<0.001) or the agreement is on another action (p<0.001). The agreed upon action is also rated higher than the same action in the other two conditions (p < 0.001 in both cases). When subjects read about a scenario where both parties have agreed to take ‘action 7’ / ‘action 70’, then the rating for taking ‘action 10’ / ‘action 100’ is significantly lower than taking ‘action 10’ / ‘action 100’ when the is no agreement or taking

'action 10' / 'action 100' when there is an agreement to take 'action 10' / 'action 100' ($p < 0.001$ and $p < 0.05$ respectively).

Taken together, this additional analysis is important because it bolsters our claim that responses we elicit in Module 1 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.

4 Additional Analysis

4.1 Comparing Lying Aversion Specifications

In Table S8 we compare three possible functional forms for lying aversion: a fixed cost of lying, a linear cost of lying and a quadratic cost of lying. We focus on the Agreement treatment (since the lying aversion model makes no prediction for the No Agreement case). It is not possible to estimate the linear cost model on the DDG data, as the lying cost is perfectly co-linear with the monetary payoff for each actions.

Table S8: Conditional logit estimation of lying aversion for the Double Dictator Game and the Bertrand Game, estimated for Agreement treatment.

	DDG		BG		
	Fixed Cost	*Quadratic Cost*	*Fixed Cost*	*Linear Cost*	*Quadratic Cost*
VARIABLES	(1)	(2)	(3)	(4)	(5)
Action Payoff	0.305***	-1.056***	0.033***	-0.009**	-0.0001
	(0.032)	(0.067)	(0.003)	(0.004)	(0.003)
Lying Aversion	-7.172***	0.206***	-4.587***	-0.036***	-0.0003***
	(0.582)	(0.013)	(0.153)	(0.003)	(0.000)
Observations	238	238	266	266	266
Log Likelihood	-314.84	-304.27	-892.40	-1139.99	-1174.77
Bayesian IC	645.43	624.29	1805.20	2300.39	2369.94

Notes: The dependent variable is the chosen action in the Double Dictator Game and the Bertrand Game; standard errors are reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Lying aversion models are only estimated for Agreement treatment.

4.2 Combining Lying Aversion with Inequality Aversion

Since Lying Aversion reverts to the selfish model in the No Promise case, and given the relatively high actions even in the case where subjects do not make a promise, it is perhaps not surprising that Lying Aversion alone does not do a particularly good job explaining our data. We therefore consider a specification that combines lying aversion with inequality aversion (as well as a specification with inequality aversion alone). The results are reported in Table S9. We find that this specification has a good overall fit, with a BIC of 2574.27 in the DDG and 4825.51 in the BG. However, in the DDG, the quality of fit seems to be driven entirely by the inequality aversion. The

model with only IA actually has a better fit than the combined model with a BIC of 2567.02. Additionally in the combined model the coefficient on Lying Aversion is not significant (albeit with a negative point estimate of -0.044). In the BG the model with IA alone has a slightly worse fit than the combined model (a BIC of 4960.77). However, both models have a substantially improved fit over the model with only lying aversion (a BIC of 5511.49), which again suggests that inequality aversion is providing most of the fit.

Table S9: Conditional logit estimation for Inequality aversion and Lying Aversion

	DDG		***BG***	
	IA	*IA+LA*	*IA*	*IA+LA*
VARIABLES	(1)	(2)	(3)	(4)
Action Payoff (β)	-0.017**	-0.007	0.080***	0.045***
	(0.008)	(0.011)	(0.004)	(0.004)
Ineq. Aversion	-0.048***	-0.047***	-0.127***	-0.091***
	(0.002)	(0.003)	(0.006)	(0.005)
Lying Aversion		-0.044		-0.042***
		(0.035)		(0.003)
Observations	620	620	620	620
Log Likelihood	-1274.68	-1273.89	-2469.34	-2396.19
Bayesian IC	2567.02	2574.27	4960.77	4825.51

Notes: The dependent variable is the chosen action in the Double Dictator Game and 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. Each observation represents a subject's choice in a particular period. For the conditional logit estimate each observation corresponds with 11 possible alternatives for the DG and 101 possible alternatives for the BG.

There is no direct reason why inequality aversion would predict different behavior between the treatments – the only reason why it can account for different action choices in the promise treatment is that the estimation is using the stated beliefs from the individual subject (leading to a different belief about the level of inequality for a given action across treatments). However, the model does not generate, and cannot explain, this difference in beliefs. The social norms model, however, both predicts such a difference in beliefs, and separately predicts a direct change in actions. In Table S10 we re-estimate the conditional logit models with 500 bootstrap replications where in each replication we randomly permute the beliefs across subjects in both treatments. In this way we can turn off the belief channel to identify the direct effect of the various preference models. In both games we now find that the Inequity Aversion models are no longer able to explain subject choices well. Instead, both of the Norms models are now the best fitting models by far. Across both games the norms component of utility remains large and positive despite the scrambled beliefs. Hence we believe that models including social norm information provide the best explanation for the effect of promises on behavior.

Table S10a: Conditional logit estimation with permuted beliefs

VARIABLES	*Selfish* (1)	*Norms* (2)	*IA* (3)	*LA* (4)	*Norms + LA* *(5)*	*IA+LA* (6)
Action Payoff (β)	0.0028	0.254***	0.008	0.043***	0.251***	0.050***
	(0.008)	(0.058)	(0.009)	(0.011)	(0.059)	(0.011)
Norms		1.450***			1.380***	
		(0.272)			(0.335)	
Ineq. Aversion			0.005**			0.005**
			(0.002)			(0.003)
Lying Aversion				-0.211***	-0.035	-0.215***
				(0.042)	(0.077)	(0.041)
Observations	620	620	620	620	620	620
Log Likelihood	-1486.60	-1370.86	-1482.81	-1455.04	-1370.191	-1450.14
Bayesian IC	2982.02	2759.37	2983.28	2927.72	2766.87	2926.77

Table S10b: Conditional logit estimation for Inequality aversion and Lying Aversion

VARIABLES	*Selfish* (1)	*Norms* (2)	*IA* (3)	*LA* (4)	*Norms + LA* *(5)*	*IA+LA* (6)
Action Payoff (β)	-0.026*	-0.006***	-0.002	-0.009***	-0.008***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.255)	(0.002)
Norms		1.072***			0.925***	
		(0.074)			(0.079)	
Ineq. Aversion			-0.005**			-0.025***
			(0.002)			(0.002)
Lying Aversion				-0.033***	-0.012***	-0.046***
				(0.004)	(0.004)	(0.003)
Observations	620	620	620	620	620	620
Log Likelihood	-2860.36	-2589.97	-2856.52	-2766.19	-2579.99	-2698.93
Bayesian IC	5731.76	5202.03	5735.13	5554.46	5193.97	5431.00

Notes: The dependent variable is the chosen action in the Double Dictator Game and Bertrand Game; standard errors are reported in parentheses with bootstrapped standard error; *** p<0.01, ** p<0.05, * p<0.1. Each observation represents a subject's choice in a particular period. For the conditional logit estimate each observation corresponds with 11 possible alternatives for the DG and 101 possible alternatives for the BG. Each estimation involves 500 bootstrap replications where the subject beliefs were permuted randomly across treatments in each replication.