Exp Econ (2010) 13: 364–377
DOI 10.1007/s10683-010-9246-4

The effects of (incentivized) belief elicitation in public goods experiments

Simon Gächter · Elke Renner

Received: 21 September 2006 / Accepted: 15 June 2010 / Published online: 13 July 2010

Abstract Belief elicitation is an important methodological issue for experimental economists. There are two generic questions: 1) Do incentives increase belief accuracy? 2) Are there interaction effects of beliefs and decisions? We investigate these questions in the case of finitely repeated public goods experiments. We find that belief accuracy is significantly higher when beliefs are incentivized. The relationship between contributions and beliefs is slightly steeper under incentives. However, we find that incentivized beliefs tend to lead to higher contribution levels than either non-incentivized beliefs or no beliefs at all. We discuss the implications of our results for the design of public good experiments.

Keywords Incentives · Beliefs · Experimental methodology · Public goods

JEL Classification C 90

Electronic supplementary material The online version of this article (doi:10.1007/s10683-010-9246-4) contains supplementary material, which is available to authorized users.

S. Gächter (✉) · E. Renner
School of Economics, Centre for Decision Research and Experimental Economics, University of Nottingham, Sir Clive Granger Building, University Park, Nottingham, NG7 2RD, UK
e-mail: simon.gaechter@nottingham.ac.uk

E. Renner
e-mail: elke.renner@nottingham.ac.uk

S. Gächter
CESifo, Munich, Germany

S. Gächter
IZA, Bonn, Germany

1 Introduction

Beliefs, like preferences and constraints, are a central concept in modern economics. Eliciting beliefs about co-players' behavior is therefore often interesting for experimental economists. Yet, when eliciting beliefs one important design question is: should subjects be financially rewarded if their stated beliefs about their opponents' choices correspond to their opponents' actual choices? Compared to a rather voluminous literature on the importance of financial incentives for making *choices*,[1] the corresponding literature for elicited *beliefs* is still rather scant.[2] This paper contributes to the literature on the role of incentives in belief elicitation.

From an experimental design point of view the question whether beliefs should be incentivized is less straightforward than whether choices should be incentivized. The reason is that there are two generic issues when beliefs are elicited. The first one is whether incentives improve results in the sense that they actually increase the accuracy of elicited beliefs and the second issue concerns possible interaction effects of (incentivized) beliefs with other variables of interest. This paper addresses both generic issues in the context of public goods experiments. We believe empirical knowledge on the role of (incentivized) beliefs in public goods experiments is important because eliciting beliefs about other group members' contributions has long been of interest to researchers.[3]

With regard to the first generic question our goal is not to assess different methods of incentivizing belief elicitation,[4] but to see whether with a given method of belief elicitation incentivizing beliefs affects the accuracy of stated beliefs. The second question is whether there are interaction effects of (incentivized) beliefs with contribution decisions. For instance, incentives for beliefs might change the *relationship* between beliefs and contributions (the focus of much research—see footnote 3) and the mere fact that beliefs are elicited at all might also affect overall contribution *levels*. For example, Croson (2000) found significantly *lower* contribution rates in her finitely repeated public goods experiments when she elicited incentivized beliefs than

[1]Camerer and Hogarth (1999) review the literature and conduct a meta-analysis (comprising 74 studies) about the impact of financial incentives for choices. For an earlier survey see Smith and Walker (1993). Hertwig and Ortmann (2001) compare practices of using incentives in psychology and economics.

[2]Some examples of studies relevant to ours are the following. Wright and Aboul-Ezz (1988) investigate the impact of incentives on the quality of frequency assessments and find that incentives increase accuracy. Erev et al. (1993) find that eliciting probability assessments can interact with other variables of interest in public goods games. Similarly, Croson (1999, 2000) look at prisoner's dilemma and public goods games and find that eliciting beliefs changes contributions compared to when beliefs are not elicited. Rutström and Wilcox (2009) look at matching pennies games and find that eliciting beliefs can change play.

[3]See Kelley and Stahelski (1970) and Kuhlman and Wimberley (1976) for early studies in psychology. In experimental economics, Offerman (1997), and Offerman et al. (1996) were among the first to elicit beliefs to understand behavior in step-level public goods. Dufwenberg et al. (2006) elicited beliefs in one-shot games to investigate the impact of framing and to evaluate theories of reciprocity and guilt aversion. Croson (2007) elicited beliefs in repeatedly played public goods games to test theories of voluntary cooperation. Gächter and Herrmann (2009) elicited beliefs to understand cross-cultural differences in cooperation behavior. Neugebauer et al. (2009) and Fischbacher and Gächter (2010) elicited beliefs to explain the declining pattern of cooperation in repeated public good games.

[4]On this question see, e.g., Holt (1986), Selten (1998), Andersen et al. (2007), Offerman et al. (2009), Palfrey and Wang (2009).

when she did not elicit beliefs. Steep incentives for accurate beliefs might help belief accuracy but might also change the whole incentive structure of the experiment (Blanco et al. 2010).

We answer our research questions with the help of three treatments: (i) a benchmark treatment where we did not elicit beliefs (called the "no beliefs treatment"); (ii) a "non-incentivized beliefs treatment" where we simply asked subjects about their estimated average contributions of other group members; and (iii) an "incentivized beliefs treatment" where subjects were paid according to the accuracy of their estimates relative to the actual average contribution of other group members.

We find that incentivizing beliefs significantly increases belief accuracy. With regard to interaction effects of belief elicitation and contributions we find that the *relationship* between beliefs and contributions is slightly stronger when beliefs are incentivized. We also find that incentivized beliefs might affect contribution *levels*: In our experiments contributions are significantly *higher* (in particular in the second half of the experiment) when elicited beliefs are incentivized but insignificantly different when beliefs are elicited but not incentivized.

Notice that the goal of this paper is not to test particular theories why beliefs and contributions might be correlated, or why beliefs might affect contribution levels. Our contribution is a basic methodological one: highlighting any tradeoff that might exist between incentivizing beliefs and possible interaction effects with contribution choices. We discuss the potential implications of our findings for designing experiments in our concluding section.

2 Design and procedures

Our design involves a standard linear public goods game, played by groups of four members. Each member has to decide on how many of 20 tokens to keep and how many tokens to contribute to a public good. The stage game payoff for each subject is given by:

$$\pi_i = 20 - g_i + 0.4 \sum_{j=1}^{4} g_j. \qquad (1)$$

From (1) it is obvious that a rational and selfish individual has an incentive to contribute nothing, whereas full contributions would be socially optimal.

We had three treatments in two of which we elicited beliefs about the other group members' average contribution. Specifically, on the same screen we prompted subjects for contribution decisions, we also prompted them for estimates how much the other three group members will contribute on average to the public good (see the instructions in the online supplementary materials for details).[5] In one treatment ("incentivized beliefs") we gave them a financial incentive for reporting beliefs accurately. We paid subjects 20 money units in every case where a participant estimated

[5] Since we asked participants to estimate the average contribution of the *other three* group members, participants had no strategic incentives to manipulate the accuracy of their estimate by changing their own contribution.

the actual contribution of others exactly right (±0.5 tokens); and 10 money units divided by the (absolute) estimation error if his or her estimate deviated by more than ±0.5 tokens from the actual contribution. In a second treatment we simply elicited beliefs without any payment for accuracy of belief estimation ("non-incentivized beliefs"). Since belief elicitation itself may change contribution behavior, we included a benchmark treatment ("no beliefs") in which we did not elicit beliefs.

We conducted all experiments at the Universities of Erfurt and Nottingham.[6] We used z-Tree (Fischbacher 2007) to conduct the experiments. Our participants were 204 undergraduates from various disciplines; 72 people participated in the "no beliefs" treatment; 68 in the "non-incentivized beliefs" treatment and 64 in the "incentivized beliefs" treatment. We allocated participants to groups randomly, but group composition remained the same throughout the ten periods of the experiment ("partners"). Thus, in total we have observations from 51 independent groups of four participants.

The participants were randomly assigned to the booths in the laboratory at the beginning of each session. The booths separated the participants visually and ensured that every individual made his or her decision anonymously and independently. The written instructions explained the game, payoffs, and procedures. Participants had to answer a set of control questions and we did not start before all participants had answered all questions correctly. In all treatments participants received their cumulative earnings according to (1). In the incentivized beliefs treatment earnings also included the belief estimation payoffs. Our experiments lasted 30 minutes on average, and participants earned €4.50 in Erfurt and £7.70 in Nottingham (these are very similar amounts in terms of purchasing power). Earnings exceed the average hourly wage of a typical student job in both locations.

3 Results

Result 1. Belief accuracy (the difference between stated beliefs and actual contributions) is significantly higher when belief elicitation is incentivized than under non-incentivized belief elicitation.

Support. A first support for Result 1 is Fig. 1. Panel A shows the distribution of the difference between estimated and actual contributions of other group members. In both treatments there is a mode at zero, that is, perfect belief accuracy. The frequency of exactly correct beliefs is 13 percent under non-incentivized beliefs and 17 percent under incentivized beliefs. When beliefs are not incentivized, 32 percent (47 percent) of beliefs differ by $\pm 1 (\pm 2)$ from others' actual contributions; this ratio is 44 percent (61 percent) under incentivized belief elicitation. The mean (median) difference is 0.90 (0.67) under non-incentivized beliefs and 0.31 (0.00) under incentivized beliefs. The standard deviation of the differences across all decisions is 4.75 under non-incentivized beliefs and 3.68 under incentivized beliefs.

[6] We used ORSEE (Greiner 2004) for recruiting participants.

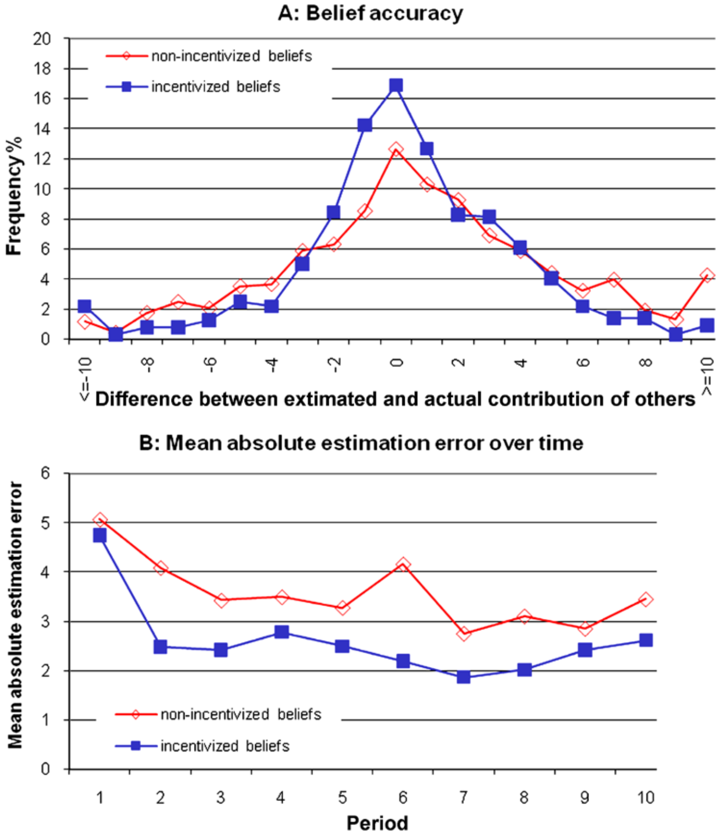
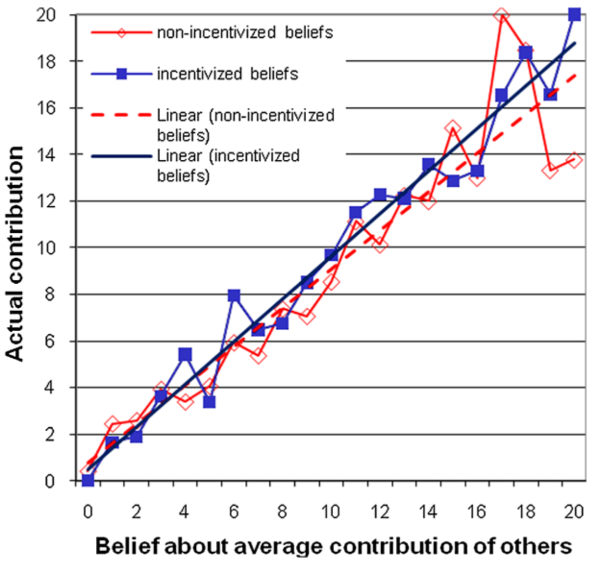


Fig. 1 Belief accuracy: distribution (panel **A**) and development over time (panel **B**)

Panel B of Fig. 1 depicts the development of the mean absolute estimation error over time. We also use this statistic for statistical testing since we are interested in any error—positive and negative. In the following we report test results based on two-sided non-parametric tests with group averages as independent observations.

We get four noteworthy results. First, there is no statistically significant difference between treatments in the mean absolute estimation error in period 1 (individual contributions as observations, two-sided Mann-Whitney test, $p = 0.406$). Second, there is a significant drop in the mean absolute estimation error from period 1 to period 2 (Wilcoxon signed-rank tests, $p = 0.008$ in non-incentivized beliefs; $p = 0.000$ under incentivized beliefs). Third, from period 2 to 10 belief accuracy stays roughly constant, in particular under incentivized beliefs (similar to Croson (2000), Fig. 2, p. 307). Finally, in each of periods 2–10 the mean absolute error is higher under non-incentivized beliefs than under incentivized beliefs. This difference is highly significant according to a two-sided Mann-Whitney test applied to the groups' average absolute estimation errors across all periods ($p = 0.00085$). A regression analysis (using a multilevel mixed random effects regression which allows for individual and

Fig. 2 The relationship between beliefs and contributions



group differences, as well as for treatment-specific residuals) confirms this result (see Table A1 in the Appendix).

Result 2. The *relationship* between beliefs and contributions is slightly steeper under incentivized than non-incentivized belief elicitation.

Support. Figure 2 provides graphical and Table A2 in the Appendix econometric support. Figure 2 shows the mean contribution for a given mean estimated contribution of the other group members. Like in previous experiments beliefs and contributions are positively correlated.[7] When beliefs are incentivized the relationship between beliefs and contributions is slightly stronger.[8]

Result 2 makes sense for the following reason (see also Rutström and Wilcox (2009) whose arguments we apply here): Suppose that estimating beliefs requires cognitive effort that is subject to random error and the random error depends on the effort subjects put into the estimation. The incentivized beliefs treatment gives subject an incentive for higher thinking effort and this should reduce the variance in the estimation error. Result 1 supports this reasoning. Since the "measurement error"

[7]There can be different reasons for such a correlation: people might project their own contributions on others (e.g., Dawes et al. 1977), or it might be an expression of reciprocity or "conditional cooperation" in general (e.g., Fischbacher et al. 2001; Fischbacher and Gächter 2010; Croson 2007; Kocher et al. 2008; Muller et al. 2008; Herrmann and Thöni 2009; Neugebauer et al. 2009; Thöni et al. 2009). Notice, however, that the purpose of the present paper is not to explain this relation but simply to see whether incentivized beliefs change this relationship, as compared to non-incentivized beliefs.

[8]This result is not robust to the estimation method, because the increase is only significant under the multilevel random effects model but not in the Tobit random effects model (see Table A2 in the Appendix). However, the Tobit model does not account for treatment-specific residuals. Since Result 1 showed that variance is lower under incentivized beliefs than under non-incentivized beliefs controlling for treatment-specific variances seems to be warranted. This argument favors the multilevel random effects model where we allow for treatment-specific residuals.

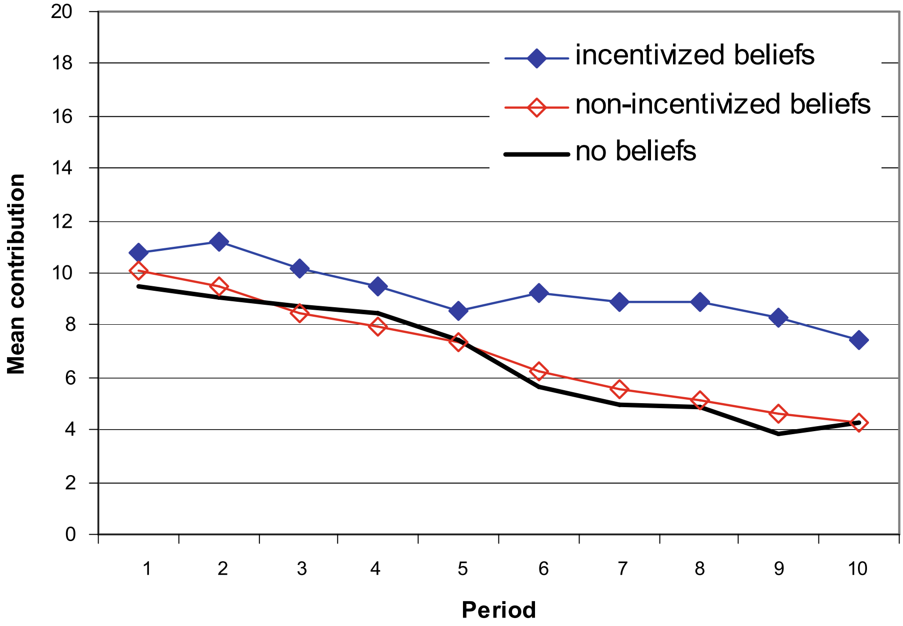


Fig. 3 Contribution levels over time

of beliefs is reduced under incentivized beliefs, the correlation between beliefs and contributions should be strengthened. This is the case empirically—hence Result 2.

Result 3. Eliciting beliefs does not change contribution *levels* relative to the no beliefs treatment if beliefs are not incentivized. By contrast, if beliefs are incentivized, we observe higher contributions in particular in the second half of the experiment.

Support. Figure 3, non-parametric tests, and regressions (Table A3 in the Appendix) provide the support for Result 3. Qualitatively, the usual picture of a declining trend emerges in all treatments (see Ledyard 1995; Herrmann et al. 2008; Fischbacher and Gächter 2010). Yet, eyeballing the time trend suggests some differences between treatments, in particular towards the end of the experiment and when beliefs are incentivized.

A statistical analysis confirms these impressions. First, a Kruskal-Wallis test with group average contributions across all periods weakly rejects the null hypothesis that the group averages are identically distributed across treatments ($p = 0.079$). Pair wise Mann-Whitney tests (using group averages across all periods) show that contributions are weakly significantly different between incentivized beliefs and non-incentivized beliefs ($p = 0.0539$); not significantly different between no beliefs and non-incentivized beliefs ($p = 0.8819$) and weakly significantly different between no beliefs and incentivized beliefs ($p = 0.05124$). Period-by-period Kruskal-Wallis tests show that significant differences only emerge after period 5, where $p \leq 0.04$ in all periods. Pair wise tests for periods 6 to 10 between no beliefs and non-incentivized beliefs reveal no significant differences (all $p > 0.58$). Pair wise comparisons between no beliefs and incentivized beliefs show significant differences at $p < 0.05$.

Result 3 is surprising given that Croson (2000) finds that contributions *decrease* when beliefs are elicited whereas Wilcox and Feltovich (2000) find that contributions are *not* affected by whether beliefs are elicited or not. Since we did not set up our experiment as a replication of neither of Croson's nor of Wilcox and Feltovich's (2000) study, any explanation of why our results differ from theirs is somewhat speculative.

One obvious possibility is differences in designs of the three studies. Our experiment is most similar to Croson's design: Many parameters were the same (groups of four participants, and ten periods in a "partner" design) or similar (in Croson's experiment subjects had an endowment of 25 tokens and the marginal per capita rate was 0.5). There is one bigger difference between our designs and Crosons's, however, and this concerns the actual level of stakes for correct beliefs.[9] To see this, recall that we asked subjects for their guess of the *average* contribution level of the other group members. Given subjects' endowment of 20 the average is therefore between 0 and 20. By contrast, Croson (2000) rewarded subjects for the accuracy of the *sum* of others' contributions, which, given subjects' endowment of 25 tokens, ranged between 0 and 75 tokens. Although the reward for a correct guess was similar in our experiment than in Croson's,[10] the difference in the range of outcomes was more than three times lower in our experiment than in hers. Therefore, the stakes were arguably much steeper in our experiment than in hers and might have induced subjects in our experiments to coordinate to garner the rewards for exact prediction. Why our subjects might have coordinated on higher contributions is an open question.

Wilcoxon and Feltovich's (2000) design is less similar to Croson's (2000) and our study: Their groups consisted of six subjects who played for six periods. Their marginal per capita return was 0.25 and the contribution decision was binary to either contribute an endowment of one token or not. Also the incentives on beliefs were implemented in a different way: One person per group was randomly selected and paid \$10 if, in a randomly selected round, he or she had correctly guessed how many of the five other group members had contributed to the public good. Given these differences it is difficult to directly compare these studies and attribute the different results to specific design features.

We also don't know whether our explanation of different incentives for correct beliefs really is the explanation for the differences in results between Croson (2000) and our study. Only new experiments, which would need to include *exact* replications of their experiments, could tell. However, for the purposes of this paper the explanation of differences in results does not matter. What matters is that our results, as well as Croson's, show that incentivized belief elicitation *can* have strong consequences, which experimentalists, who design new experiments, should be aware of. We discuss potential implications for the design of experiments in the next section.

[9] We are grateful to a referee who pointed out this possibility.

[10] In both experiments a correct guess (+/ − .5 points in our experiment) earns a subject an amount equivalent to the endowment in each period (25 points in Croson's experiment, 20 points in our experiment) and half that amount divided by the (absolute) estimation error is paid for all other guesses.

4 Concluding discussion

In our view, the implications of our results for the design of public goods experiments depend on the actual research question. If beliefs are the focus of interest, then our Result 1 suggests that beliefs should be incentivized because belief accuracy is higher when beliefs are incentivized than when they are hypothetical. If the researcher is mainly interested in the *relationship* between beliefs and contributions, then our Result 2 suggests that incentives on beliefs do strengthen this relationship, although the effect is quite small. If the researcher is afraid that belief elicitation leads to behavioral results that he or she would not obtain when not asking for beliefs, then our Result 3 suggests that belief elicitation should not be incentivized, because without incentives for correct beliefs we did not get a significant difference between the no belief treatment and the non-incentivized beliefs treatment. If beliefs are incentivized, interaction effects of beliefs and contributions are likely to exist and can, given our results and those of Croson (2000) and Wilcox and Feltovich (2000), apparently go in either direction.

Whatever the explanation for these conflicting findings is, one point seems clear and was already pointed out by Croson (2000), p. 312: "If the act of eliciting beliefs affects behavior, experimentalists need to think carefully about their procedures and the information collected in their experiments. In particular, they need to examine and take into account the possible effects of their elicitation procedures on behavior." Our results firmly support this conclusion. In particular, Results 1 and 3 suggest that, depending on the research question, the researcher may face a tradeoff: high incentives for exact beliefs increase belief accuracy but strongly incentivized beliefs may also bias contributions away from levels that would be observed were beliefs not elicited at all or only hypothetically.

We are of course aware that, in a strict sense, our conclusions only have validity in the context of public good experiments. However, (i) given the importance of beliefs in theoretical models, (ii) an increased attention to beliefs by empirical economists (e.g., Manski 2004) and behavioral game theorists (e.g., Dufwenberg and Gneezy 2000; Nyarko and Schotter 2002; Bhatt and Camerer 2005; Costa-Gomes and Weizsäcker 2008) and (iii) the findings of Croson (2000) and Wilcox and Feltovich (2000) and our new results, we think it is a methodologically important task to investigate the impact of (incentivized) belief elicitation in other economically interesting tasks as well.

Acknowledgements We gratefully acknowledge financial support from the University of Erfurt. We thank Tim Cason, Eva Poen, Martin Sefton, Christian Thöni and in particular two anonymous referees for their very helpful comments.

Appendix: Regression analyses

In the regression analyses we used two approaches. First, a multilevel mixed random effects regression which allows for individual and group differences, as well as for

Table A1 Mean absolute estimation error under incentivized and non-incentivized belief elicitation

Mean absolute estimation error (absolute difference between stated belief and actual average contribution of others)		
	Multilevel mixed random effects	Random effects Tobit
Dummy Incentivized Beliefs (IB)	-1.430^{***}	-1.505^{***}
	(0.484)	(0.464)
Period	-0.093^{***}	-0.121^{***}
	(0.047)	(0.044)
Period × IB	0.066	0.077
	(0.059)	(0.063)
First period	1.203^{***}	1.183^{***}
	(0.395)	(0.418)
First period × IB	1.043^{*}	1.107^{*}
	(0.563)	(0.600)
Constant	3.959^{***}	4.007^{***}
	(0.367)	(0.323)
σ(group)	0.684^{***}	
	(0.171)	
σ(subject)	1.002^{***}	1.239^{**}
	(0.119)	(0.116)
σ(residual)		2.779^{***}
		(0.060)
σ(residual non-incentivized beliefs treatment)	2.977^{***}	
	(0.085)	
σ(residual incentivized beliefs treatment)	2.219^{***}	
	(0.065)	
Observations	1320	1320
Wald chi2(5)	98.70^{***}	93.72^{***}

Standard errors in parentheses; $^{*}p < 10\%$; $^{**}p < 5\%$; $^{***}p < 1\%$

treatment-specific residuals, and second a random effects Tobit estimation (controlled for individual differences) to account for the fact that the data are censored.[11]

Table A1: Mean absolute estimation error (Result 1)

Table A1 reports the results of a regression where the dependent variable is the mean absolute estimation error, and the independent variables are a dummy for the incentivized beliefs treatment, the period index, a dummy for the first period, and two interaction variables to control for treatment-specific differences. The results show that the mean absolute estimation error is highly significantly lower under incentivized

[11] We used the software package Stata 11 for estimation. We estimated the multilevel random effects model using the `xtmixed` command and for the random effects Tobit we used the `xttobit` command.

Table A2 The correlation between contributions and beliefs

	Contributions	
	Multilevel mixed random effects#	Random effects Tobit
Belief about average contribution of others	0.660***	0.986***
	(0.045)	(0.064)
Belief × Dummy incentivized beliefs treatment	0.134**	0.084
	(0.066)	(0.098)
Dummy incentivized beliefs treatment	−1.003	−0.688
	(1.067)	(1.546)
Period	−0.232***	−0.396***
	(0.062)	(0.083)
Period × Dummy incentivized beliefs treatment	0.167**	0.268**
	(0.081)	(0.112)
Constant	3.034***	0.064
	(0.747)	(1.040)
σ (subject)	3.413***	5.060***
	(0.240)	(0.376)
σ (residual)		4.925***
		(0.130)
σ (residual non-incentivized beliefs treatment)	4.089***	
	(0.117)	
σ (residual incentivized beliefs treatment)	3.484***	
	(0.103)	
Observations	1320	1320
Wald chi2(5)	704.53	621.39

#Controlling for both group and individual differences eliminated the group level. In this estimation we therefore only controlled for individual differences

Standard errors in parentheses; $^{*}p < 10\%$; $^{**}p < 5\%$; $^{***}p < 1\%$

beliefs than under hypothetical beliefs (the benchmark). The estimation errors are reduced over time and are significantly higher in the first period compared to the rest. All these observations are robust to the estimation method (multilevel mixed random effects with treatment-specific residuals, or random effects Tobit).

Table A2: Correlation between contributions and beliefs (Result 2)

We regress contributions on elicited beliefs, the period index, and interaction variables with the incentivized beliefs treatment. The random effects regression finds that under incentivized beliefs contributions appear to be significantly higher than under non-incentivized beliefs (the benchmark). However, this result is not robust to the estimation method, because in a random effects Tobit regression the coefficient on "Belief × Dummy incentivized beliefs treatment" loses its significance. Due to substantial left and right censoring of the data (32.6 percent of all observations are 0

Table A3 Contributions as a function of treatment and time

	Contribution	
	Multilevel mixed random effects	Random effects Tobit
Dummy non-incentivized beliefs treatment	0.108	0.049
	(1.262)	(1.305)
Dummy incentivized beliefs treatment	0.716	0.551
	(1.269)	(1.314)
Period	-0.698^{***}	-1.003^{***}
	(0.065)	(0.087)
Period×Dummy non-incentivized beliefs treatment	0.025	−0.036
	(0.089)	(0.127)
Period×Dummy incentivized beliefs treatment	0.346^{***}	0.506^{***}
	(0.086)	(0.124)
Constant	10.509^{***}	11.138^{***}
	(0.887)	(0.903)
σ (group)	2.939^{***}	
	(0.412)	
σ (subject)	3.227^{***}	6.234^{***}
	(0.221)	(0.367)
σ (residual)		6.205^{***}
		(0.133)
σ (residual no beliefs treatment)	4.999^{***}	
	(0.138)	
σ (residual non-incentivized beliefs treatment)	4.604^{***}	
	(0.132)	
σ (residual incentivized beliefs treatment)	4.073^{***}	
	(0.120)	
Observations	2040	2040
Wald chi2(5)	281.18^{***}	297.55^{***}

Standard errors in parentheses; $^{*}p < 10\%$; $^{**}p < 5\%$; $^{***}p < 1\%$

or 20) the Tobit approach might be considered more appropriate. See, however, our comments in footnote 8.

Table A3: Contributions as a function of treatment and time (Result 3)

Table A3 reports regression results that relate contributions to treatment dummies, the period index, and two interaction variables of period and treatment, to detect treatment-specific differences in the temporal development of contributions. The results show that contributions are significantly higher over time in the incentivized beliefs treatment, compared to the control treatment. By contrast, contributions in the non-incentivized treatments do not develop significantly differently from the control treatment. These findings are robust to the estimation technique (multilevel mixed random effect, or random effects Tobit).

References

Andersen, S., Fountain, J., Harrison, G. W., & Rutström, E. E. (2007). *Eliciting beliefs: theory and experiments*. Working Paper 07-08, Department of Economics, College of Business Administration, University of Central Florida.

Bhatt, M., & Camerer, C. F. (2005). Self-referential thinking and equilibrium as states of mind in games: fMRI evidence. *Games and Economic Behavior*, *52*, 424–459.

Blanco, M., Engelmann, D., Koch, A. K., & Normann, H.-T. (2010, forthcoming). Belief elicitation in experiments: Is there a hedging problem?. *Experimental Economics*. 10.1007/s10683-010-9249-1.

Camerer, C. F. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, *19*, 7–42.

Costa-Gomes, M. A., & Weizsäcker, G. (2008). Stated beliefs and play in normal-form games. *Review of Economic Studies*, *75*, 729–762.

Croson, R. (1999). The disjunction effect and reason-based choice in games. *Organizational Behavior and Human Decision Processes*, *80*, 118–133.

Croson, R. (2000). Thinking like a game theorist: Factors affecting the frequency of equilibrium play. *Journal of Economic Behavior & Organization*, *41*, 299–314.

Croson, R. (2007). Theories of commitment, altruism and reciprocity: Evidence from linear public goods games. *Economic Inquiry*, *45*, 199–216.

Dawes, R. M., McTavish, J., & Shaklee, H. (1977). Behavior, communication, and assumptions about other peoples behavior in a commons dilemma situation. *Journal of Personality and Social Psychology*, *35*, 1–11.

Dufwenberg, M., Gächter, S., & Hennig-Schmidt, H. (2006). *The framing of games and the psychology of strategic choice*. CeDEx Discussion Paper 2006-20, University of Nottingham.

Dufwenberg, M., & Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior*, *30*, 163–182.

Erev, I., Bornstein, G., & Wallsten, T. S. (1993). The negative effects of probability assessments on decision quality. *Organizational Behavior and Human Decision Processes*, *55*, 78–93.

Fischbacher, U. (2007). Z-tree: Zurich toolbox for readymade economic experiments. *Experimental Economics*, *10*, 171–178.

Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public good experiments. *American Economic Review*, *100*, 541–556.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*, 397–404.

Gächter, S., & Herrmann, B. (2009). Reciprocity, culture, and human cooperation: Previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society B – Biological Sciences*, *364*, 791–806.

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer & V. Macho (Eds.), *Forschung und Wissenschaftliches Rechnen. GWDG Bericht 63*. Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.

Herrmann, B., & Thöni, C. (2009). Measuring conditional cooperation: A replication study in Russia. *Experimental Economics*, *12*, 87–92.

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, *319*, 1362–1367.

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: a methodological challenge for psychologists?. *Behavioral and Brain Sciences*, *24*, 383.

Holt, C. A. (1986). Scoring-rule procedures for eliciting subjective probability and utility functions. In P. Goel & A. Zellner (Eds.), *Bayesian inference and decision techniques*. Amsterdam: Elsevier.

Kelley, H., & Stahelski, A. (1970). Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology*, *16*, 190–219.

Kocher, M. G., Cherry, T., Kroll, S., Netzer, R. J., & Sutter, M. (2008). Conditional cooperation on three continents. *Economics Letters*, *101*, 175–178.

Kuhlman, D. M., & Wimberley, D. L. (1976). Expectations of choice behavior held by cooperators, competitors, and individualists across four classes of experimental games. *Journal of Personality & Social Psychology*, *34*, 69–81.

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In A. E. Stapel Roth & J. H. Kagel (Eds.), *The handbook of experimental economics*. Princeton: Princeton University Press.

Manski, C. F. (2004). Measuring expectations. *Econometrica*, *72*, 1329–1376.

Muller, L., Sefton, M., Steinberg, R., & Vesterlund, L. (2008). Strategic behavior and learning in repeated voluntary-contribution experiments. *Journal of Economic Behavior & Organization*, *67*, 782–793.

Neugebauer, T., Perote, J., Schmidt, U., & Loos, M. (2009). Self-biased conditional cooperation: On the decline of cooperation in repeated public goods experiments. *Journal of Economic Psychology*, *30*, 52–60.

Nyarko, Y., & Schotter, A. (2002). An experimental study of belief learning using elicited beliefs. *Econometrica*, *70*, 971–1005.

Offerman, T. (1997). *Beliefs and decision rules in public good games. Theory and experiments*. Dordrecht: Kluwer.

Offerman, T., Sonnemans, J., & Schram, A. (1996). Value orientations, expectations and voluntary contributions in public goods. *Economic Journal*, *106*, 817–845.

Offerman, T., Sonnemans, J., Van De Kuilen, G., & Wakker, P. P. (2009). A truth serum for non-bayesians: Correcting proper scoring rules for risk attitudes. *Review of Economic Studies*.

Palfrey, T. R., & Wang, S. W. (2009). On eliciting beliefs in strategic games. *Journal of Economic Behavior & Organization*, *71*, 98–109.

Rutström, E. E., & Wilcox, N. T. (2009). Stated beliefs versus inferred beliefs: A methodological inquiry and experimental test. *Games and Economic Behavior*, *67*, 616–632.

Selten, R. (1998). Axiomatic characterization of the quadratic scoring rule. *Experimental Economics*, *1*, 43–62.

Smith, V. L., & Walker, J. M. (1993). Monetary rewards and decision cost in experimental economics. *Economic Inquiry*, *31*, 245–261.

Thöni, C., Tyran, J.-R., & Wengström, E. (2009). *Microfoundations of social capital*. Discussion Paper No. 09-24, Department of Economics, University of Copenhagen.

Wilcox, N. T., & Feltovich, N. (2000). *Thinking like a game theorist: Comment*. Mimeo, University of Houston, available at http://www.uh.edu/~nfelt/papers/belief.pdf.

Wright, W. F., & Aboul-Ezz, M. E. (1988). Effects of extrinsic incentives on the quality of frequency assessments. *Organizational Behavior and Human Decision Processes*, *41*, 143–152.