Costa-Gomes, Miguel A.; Huck, Steffen; Weizsäcker, Georg

Beliefs and actions in the trust game: Creating instrumental variables to estimate the causal effect

Games and Economic Behavior

Games and Economic Behavior 88 (2014) 298–309

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Beliefs and actions in the trust game: Creating instrumental variables to estimate the causal effect ☆

Miguel A. Costa-Gomes [a], Steffen Huck [b,*], Georg Weizsäcker [c]

[a] University of St Andrews, School of Economics & Finance, Castlecliffe, The Scores, Fife KY16 9AR, United Kingdom
[b] WZB Berlin Social Science Center, Reichpietschufer 50, 10785 Berlin, Germany
[c] Humboldt University Berlin, School of Business and Economics, Spandauer Str. 1, 10178 Berlin, Germany

ARTICLE INFO

Article history:
Received 8 August 2013
Available online 22 October 2014

JEL classification:
C72
C81
C91
D84

Keywords:
Social capital
Trust game
Instrumental variables
Belief elicitation

ABSTRACT

In many economic contexts, an elusive variable of interest is the agent's belief about relevant events, e.g. about other agents' behavior. A growing number of surveys and experiments asks participants to state beliefs explicitly but little is known about the causal relation between beliefs and actions. This paper discusses the possibility of creating exogenous instrumental variables for belief statements, by informing the agent about exogenous manipulations of the relevant events. We conduct trust game experiments where the amount sent back by the second player (trustee) is exogenously varied. The procedure allows detecting causal links from beliefs to actions under plausible assumptions. The IV-estimated effect is significant, confirming the causal role of beliefs.

1. Introduction

In subjective expected utility theory and related models, the agent's expectations can be viewed as a pure *as-if* construct, meaning that expectations are no more than an elegant way of summarizing choice data. Choice is represented by a hypothetical function of expectations—for example, the expected utility function—and choice is thus the fundamental concept. Any assumption that one may make about expectations is really an assumption about the nature of choice. A much more literal interpretation of expectations is that they are *real*, meaning that they are independent entities that have some physical incarnation and that can in principle be accessed directly, for example by asking people to state them. Much can be said in favor of such a literal interpretation of expectations, not least that humans are able to express expectations even about variables that are irrelevant for their choices. But if expectations are independent entities, one should be able to influence them and thereby measure their effects on choice. This leads to the straightforward empirical question whether choices are driven by beliefs. We address this question in the context of trust games.

☆ We thank Ramses Abul Naga, Orazio Attanasio, Charles Bellemare, Jürgen Bracht, Jörg Breitung, Christoph Breunig, Samuele Centorrino, Syngjoo Choi, Brendan Kline, Costas Meghir, Lars Nesheim, Thomas Siedler and audiences at Alicante, Autonoma de Barcelona, CEU Budapest, City University, DIW Berlin, Exeter, Glasgow, Innsbruck, Jena, Lausanne, Paris I, Royal Holloway, UCL and WZB for their comments. We are grateful for financial support from the U.K. Economic and Social Research Council (ESRC-RES-1973), the European Research Council (ERC-263412) and the ELSE centre at UCL (ESRC-RES-538-28-1001). The experimental sessions were conducted with the excellent support of Rong Fu, Tom Rutter, Brian Wallace and Mark Wilson.
* Corresponding author.
E-mail addresses: miguel.costa-gomes@st-andrews.ac.uk (M.A. Costa-Gomes), huck@wzb.eu (S. Huck), weizsaecker@hu-berlin.de (G. Weizsäcker).

http://dx.doi.org/10.1016/j.geb.2014.10.006
0899-8256/

The question of causality of beliefs has important consequences for policy interventions. Many policy campaigns target beliefs—e.g. asserting the reality of climate change, or bolstering investor confidence—to bring about behavioral change. But empirically, the role of beliefs needs to be examined. Researchers have increasingly turned to belief elicitation procedures where the agents state their expectations explicitly. Trust game experiments (following Berg et al., 1995) provide a frequent context for such methods. Fehr et al. (2003), Bellemare and Kröger (2007), Sapienza et al. (2013) and Naef and Schupp (2009), among others, ask their experimental participants to state expectations on how much money other participants will return if trusted. They find a strong correlation as well as much explanatory power when regressing the level of trust on stated expectations. Yet it remains unanswered whether the variance in trust arises *because of* the variance in stated beliefs or whether the co-variation in the two variables is driven by other, omitted variables that capture unobservable differences between the participants. For example, participants who are likely to trust others may also be relatively trustworthy and may project their own type onto others. The player's type would be an omitted variable that creates an endogeneity problem for the data analyst. A natural reason for such a type-driven correlation between beliefs and actions is the perception of social norms. Among the experimental participants who are assigned the role of trustors, presumably some view a high investment in the game as the "right" thing to do, given that it maximizes social surplus. The social norm's perception may depend on unobservable factors like the participants' education, cultural influences or even the framing employed in the experiment. These unobservables likely influence both beliefs *and* actions. The same participant who invests a large amount may thus also predict that the participant in the other player role will return a large amount because this, too, is arguably the "right" thing to do. A positive correlation between beliefs and actions would arise—without implying anything about a causal influence of one variable on the other.[1]

Such a correlation is not necessarily a "behavioral" phenomenon but can arise as an equilibrium outcome of a natural game of incomplete information. We develop a simple illustration of this in Appendix A. Players interact in a mini trust game with just two actions for each player: whether to trust or not, and whether to reciprocate or not. Both players are aware of the social norm that prescribes trust and its reciprocation but there prevails some uncertainty about whether deviations from the norm will be sanctioned. Players receive correlated signals about the likelihood of sanctions. Appendix A shows that even with relatively little correlation between the players' signals the Bayesian Nash equilibrium involves a strong correlation between the trustor's own action and her belief about the opponent's action. The driver of both variables is the trustor's perception of the likelihood of sanctions (a variable that is omitted in most empirical analyses). The example also shows that despite the strong correlation between the trustor's belief and action, an exogenous shift of the trustor's belief about the opponent's action has a relatively small effect on her action. It would therefore be misleading to interpret the strong correlation between beliefs and actions as evidence that one drives the other.

This example only suggests one particular omission in the analyst's model—yet many other omitted variables apart from social norms might have an effect on actions and beliefs. The example's message is merely that the players may well have good reasons (here, play an equilibrium in a larger game) to exhibit correlations between beliefs and actions that the researcher may mis-interpret as a causal relation. To measure the effect of a belief change on actions, one needs more powerful observations than simple correlations.

In Section 2, we describe a technique to measure the effect in the context of a trust game, involving the artificial creation of an instrumental variable. The creation of instrumental variables in the laboratory is a technique employed previously by Ham et al. (2005) and Gill and Prowse (2014)—they measure the causal role of endogenous variables other than beliefs and for different dependent variables. Our game is a simultaneous-move version of Berg et al. (1995) trust game and the instrument is a zero-mean random shift that exogenously increases or reduces the trustee's level of re-payment. The realization of the random shift is known to the trustor, thus affecting her belief about the final level of re-payment and potentially affecting her action. The trustee is informed of the existence of the shift and of its distribution. However, she is not informed about the realization of the shift, and her behavior remains unaffected by the realization.[2] The trustor's belief about the trustee's behavior (her chosen level of re-payment prior to its manipulation through the shift) should therefore also be unaffected by the realization of the shift. Our data confirm these predictions. At the same time, the beliefs about the payoff-relevant event—the level of re-payment including the shift—react strongly to the exogenous variation, which is necessary to apply an IV estimation. Regarding the "exclusion restriction" requirement of IV, that the instrument influences the actions only via the beliefs about the level of re-payment, we argue that it is natural to make this assumption because the instrument is an element of the statistic that the belief is formed about (the level of re-payment), and does not enter the interaction in any other way.

To check for the validity of the design, the trust game is played under two different conditions—with and without the instrument. The no-instrument condition is a control that serves two key purposes: it allows checking whether the introduction of the instrumentation technique has any undesirable influences on the data generating process and it generates the benchmark "naive" estimate of the connection between beliefs and actions. Consistent with the previous literature, we find a strong correlation between the two variables. Crucially, the IV results indicate a causal link between beliefs and

[1] A downward bias may arise due to measurement error in the explanatory variable. Under classical errors-in-variables assumptions, our instrumental variable would address this issue but we do not pursue this argument further.

[2] An experimental procedure that is related to ours is to replace one player's choice by an exogenous random move as has been done in several experimental studies on reciprocity. See, in the context of the trust game, Cox (2004) among others. In short, these studies mainly differ from ours because they replace the trustor's action by a random move, whereas we manipulate the trustee's move.

actions. Exogenous belief variation has a strong and significant impact on choices. The average marginal (proportional) effect of beliefs on actions is 0.5 which is insignificantly smaller than the non-instrumented analysis suggests.

These findings constitute, to our knowledge, the first evidence supporting that first-order beliefs drive actions in the trust game. From a methodological point of view, our paper emphasizes the issue of causality in belief elicitation studies. Causal links between beliefs and actions were implicitly suggested not only in experiments with belief elicitation (McKelvey and Page, 1990; Offerman et al., 1996; Croson, 2000; Huck and Weizsäcker, 2002; Nyarko and Schotter, 2002, and many later studies) but also in survey studies that use stated expectations about relevant market variables (see Manski, 2004, and Attanasio, 2009, for useful overviews). Both literatures contain rich sets of observations that are consistent with a causal influence of beliefs on actions, but the endogeneity of beliefs and actions is rarely addressed in the analyses.[3, 4]

2. Experimental design: instrumental variables for belief statements

Our experimental design revolves around a continuous trust game with two players. We study two versions of this game, the game with instrument (Condition I) and the game without instrument (Condition NI). In addition to the choice data we collect the trustors' beliefs about the actions played by the trustees.

Condition NI serves as a control. First, it allows for a validity check of the instrument: whether or not it affects behavior in undesirable ways. In field studies that involve IV methods, this is less of a concern as the instrument is usually part of the natural decision making environment. But with an artificial instrument we must check that the instrumentation technique is neutral in the sense that its presence alone does not distort the data generating process.

Second, and no less important, Condition I provides the main comparison benchmark for our IV results: it is the usual laboratory environment for trust games. Condition I's non-instrumented estimates will also be reported but are only partially relevant because the instrumentation generates additional variance in beliefs. Under the hypothesis that an omitted variable is at work, a non-instrumented analysis on the data from Condition I yields a biased (attenuated) estimate of the relationship of interest.

We note that in all experimental sessions subjects also played a second type of trust game with binary actions. This game, too, was played in a variant with and a variant without an instrument. However, as documented extensively in the paper's previous version (Costa-Gomes et al. 2010), the instrument employed in the binary trust game failed our tests for invasiveness and hence we focus here on the continuous trust game.[5]

For the collection of belief statements, we employ a quadratic scoring rule that is incentive compatible in the sense of theoretically eliciting the mean of the subjectively expected distribution, under the assumption that subjects are risk neutral.[6]

We conducted our experimental sessions at University College London and at the University of York, with a roughly equal number of subjects in each treatment at each location, as reported in Table 1. In all, 434 experimental subjects participated in our sessions. Subjects earn points by playing two games and one belief elicitation task, which are then converted into money at an exchange rate of 2.5 pence per point, resulting in an average variable payment of £13.12.[7] Sessions lasted about 90 minutes from the moment the subjects were seated until leaving the laboratory after collecting their payments.

[3] Notable exceptions are the papers by Bellemare, Kröger and van Soest (2008, 2011a) and Bellemare et al. (2011) who estimate structural econometric models that include covariance between beliefs and actions, and Smith (2013) who studies experimental public goods games and uses lagged actions by the opponent as instruments for stated beliefs. Bellemare, Kröger and van Soest study first-order beliefs of proposers (2008) and responders (2011a) in the ultimatum game. Bellemare et al. (2011) study second-order beliefs in a sequential game akin to the trust game. Their structural models allow the parameters of an agent's other-regarding preference to be jointly determined with beliefs—an endogeneity that is confirmed in the data. Smith's (2013) IV regressions in public goods games also point at a substantial endogeneity of actions and beliefs. Our results indicate only a milder endogeneity problem in the trust game.

[4] A further important set of close relatives to our paper are field experiments that vary informational conditions in different economic contexts, see e.g. Jensen (2010) and Dupas (2011). Another related literature is summarized in Guiso et al. (2006, 2009) showing evidence of a causal role of culture on both actions and beliefs.

[5] The instrument used there is different from the instrument used in the continuous trust game and causes undesirable biases in beliefs. Specifically, 25% of elicited beliefs in the binary game are justifiable only by prior beliefs that lie outside the probability simplex. In the paper's earlier version we also carefully examine whether the continuous game data can be analyzed separately from the binary data. The experimental design involves four types of trust games (either continuous or binary and either with or without an instrument) and the protocol was such that each subject played just two games (without feedback) ensuring that she would play once in a binary and once in a continuous game, once with and once without an instrument, and once as a trustor and once as a trustee. We detected no spillover from the binary game onto the continuous trust game data.

[6] The quadratic scoring rule has been used by numerous studies. Belief elicitation procedures usually do not generate large distortions of choice data (see, e.g., Costa-Gomes and Weizsäcker, 2008) but some studies contain evidence on intrusive effects (Croson, 2000, and Rutström and Wilcox, 2009; Gächter and Renner, 2010). In our experiment, the danger of such an intrusion appears small, not least as we elicit beliefs *after* the choices. Nevertheless we readily admit that our method of payment could be cleaner as our subjects could in principle use their belief statements to hedge their positions. But such behavior would require considerable sophistication and risk aversion; existing evidence from controlled experiments (Blanco et al., 2010; Armantier and Treich, 2013) does not produce clear evidence that subjects rationally hedge in belief elicitation tasks. For recent discussions of belief elicitation methods see Armantier and Treich (2013) and Trautmann and van de Kuilen (2014).

[7] In addition, there was a show-up fee of £5 at UCL and £4 at York, chosen in each case so as to coincide with the show-up fee of a different experiment being run at the respective lab at the same time.

Table 1
Overview of experimental conditions.

Condition	# York subjects	# UCL subjects	# Total subjects
I	124	120	244
NI	94	96	190

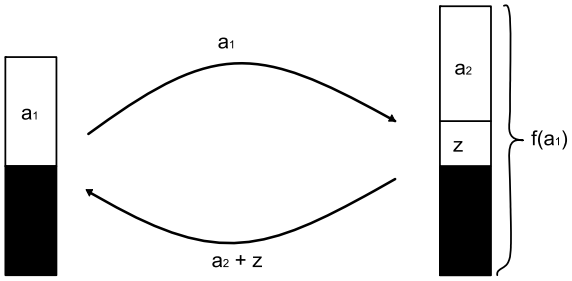

Fig. 1. Illustration of the continuous trust game with instrument. Player 2 knows only the distribution of z and chooses action a_2. Player 1 knows the distribution of z and the value of z before choosing action a_1 and belief statement b_1. $f(a_1)$ indicates that player 2's account balance depends on a_1.

2.1. The continuous trust game and the shift instrument

Each of two players initially receives an "account" that contains 100 points. The trustor, here labeled "participant X", chooses the share a_1 of her points that are to be transferred to the trustee, "participant Y". The transfer is productive—every point that the trustor sends is tripled on the way to the trustee. Simultaneously, i.e. without knowing the trustor's transfer, the trustee decides how much to transfer back from the total that she has in her account after X's transfer. The trustee, like the trustor, makes a decision about a relative "transfer share" a_2, not an absolute amount.

The transfer shares a_1 and a_2 are restricted to lie in the interval $[0.2, 0.8]$. Thus the trustor can transfer between 20 and 80 points, which are tripled and added to the trustee's amount, resulting in an account balance for the trustee between 160 and 340 points. Of these points, the trustee can transfer back a share of between 0.2 and 0.8 but has to do so without knowing the exact balance in her account.

The instrumental variable is a shift z that increases or decreases the trustee's transfer share by a value between -0.2 and 0.2, drawn from a uniform probability distribution over the 41 values on the grid $\{-0.2, -0.19, ..., 0, ..., 0.19, 0.2\}$. Both participants are informed that the trustee's transfer share a_2 is added to the zero-mean random variable z. The trustor is, in addition, informed about the realized value of z, while the trustee is not.

For example, suppose that upon being informed that the realization of the shift z is 0.05, the trustor transfers a share $a_1 = 0.5$ of her initial balance of 100 points. This would lead to intermediate account balances of 50 and 250 points for the trustor and trustee, respectively. Suppose further that the trustee decides to transfer $a_2 = 0.25$. Hence, the actual transfer to the trustor would be a share of 0.3 $(= 0.25 + 0.05)$ of the trustee's intermediate balance, leading to final balances of 125 and 175 points for the trustor and trustee, respectively. The game's rules are illustrated in Fig. 1.

We explained the shifter z to participants as follows (for full instructions, see online Appendix C):

> "There is one important detail about the transfer out of Participant Y's account. The computer adjusts the share that is actually transferred from Participant Y's account to Participant X's account. More specifically, the computer will adjust Y's transfer share in a random way, increasing or reducing it by up to 20 percentage points. That is, the computer will generate a number that we call "CHANGE TO Y's TRANSFER SHARE" by picking a random percentage number among $-20\%, -19\%, \ldots, 0\%, \ldots, +19\%, +20\%$. Each of the whole-numbered percentages in this range is equally likely."

The instructions continue by giving a further illustration of the instrumental variable and its effects on payoffs.

After making their choices, the trustor is asked to report her belief statement about the trustee's "adjusted transfer share", i.e. about the sum $\widetilde{a}_2 = a_2 + z$. The belief statement is rewarded according to the quadratic scoring rule

$$\pi_b = A - c(\widetilde{a}_2 - b_1)^2,$$

where b_1 is Participant X's belief statement about Participant Y's transfer share, and the parameter values are $A = c =$ 250 points.[8] This elicitation procedure applies both when the game is played with and without the instrument—when

[8] Under the assumption of risk neutral subjective expected utility the agent maximizes $A - c\int(\widetilde{a}_2 - b_1)^2 dF(\widetilde{a}_2)$ where F is the subjective distribution function of random variable $\widetilde{a}_2$. Reporting the mean of the distribution is optimal, as indicated by the first-order condition, $-2c\int(\widetilde{a}_2 - b_1)dF(\widetilde{a}_2) = 0 \Longleftrightarrow b_1 = \int \widetilde{a}_2 dF(\widetilde{a}_2)$.

played without the instrument, the trustor is simply asked about the trustee's transfer a_2. At the time when participants choose the actions in the game, none of them is made aware of the subsequent belief elicitation task.

Importantly, the instrument z is generated independently of all other relevant random variables. This property justifies the exogeneity assumption required for IV. Consider the bivariate linear projection of the trustor's transfer share a_1 on her stated beliefs b_1:

$$a_1 = \beta_0 + \beta_1 b_1 + u \tag{1}$$

The "exclusion restriction" for OLS requires that while the error term u is potentially confounded with b_1 e.g. due to omitted variables, the instrumental variable z is orthogonal to u. (In the regressions of the next section, we use Tobit instead of OLS models, but the same logic applies for Tobit. See e.g. Angrist and Pischke (2009) for a wider discussion of exclusion restrictions.[9]) The exclusion restriction is key for the causal inference—indeed we designed the experiment to make it maximally plausible. Since z is independently generated in the laboratory, we can rule out that u has an influence on z, or that any omitted variable may co-determine u and z. It remains an assertion that z does not influence u. We regard this as a reasonable assertion because z is a summand of $\widetilde{a}_2$, which is the statistic that beliefs b_1 are formed about, and because z does not elsewhere enter the interaction. Section 3 will contain results that demonstrate that belief statements are indeed strongly responsive to z. In fact, participants respond in a way that is fully consistent with the hypothesis that they simply add z to their beliefs about a_2.

We insert a note of caution about the simultaneous-move game: the trust game is usually played as a sequential game, where the trustor observes the trustee's action before her own move. In such a sequential format reciprocal motives can influence the trustee's decision process (see e.g. the results by Servatka et al., 2008). In our game, the anticipation of reciprocal behavior is impossible. More generally, using the simultaneous trust game simplifies the trustor's belief about the trustee's transfer. (In the sequential version the trustor's beliefs would specify such a distribution for each possible action of her own.)

3. Results

3.1. Preliminaries: data pooling, descriptive summary and checks for invasiveness of the instrument procedure

Data pooling. We first determine whether there are any statistically significant differences between the data collected at UCL and at York, and whether there are any order effects on either actions or stated beliefs (recall that each participant played two versions of the trust game). The absence of major differences allows us to pool the data and simplify the subsequent analysis.

Initially, we pair the two treatments in which the game was played under the same instrument condition at each of the locations, thereby testing for order effects. The absence of such order effects leads us to pool the data and test for laboratory effects, by comparing the data collected at the two different locations. We apply Kolmogorov–Smirnov's two-sample exact test to both players' transfer shares and to the trustor's belief statements and find no statistically significant order or laboratory effects, for any of the player roles or for any instrument condition.[10] Therefore, in the subsequent data analysis we use the pooled data played under each instrument condition.

Data summary and checks for invasiveness. As part of the data summary, we examine whether the presence of an instrument has undesired effects on how subjects play the games. More specifically, we check that the mere introduction of the instrument does not affect the behavioral variables except through the channel of influencing the beliefs. We focus on the trustor's data as the trustee's role in this study is accessory and only serves the purpose of generating an uncertain re-payment.

Our first step is to use the beliefs stated under Condition I (the beliefs about the trustee's "adjusted" transfer share after manipulation through the instrument) to construct the *underlying* beliefs about the behavior of the human opponent. We then check whether these inferred underlying beliefs are "admissible", i.e., whether one could hold such beliefs about the trustee's transfer share. More concretely, suppose that upon being informed that the shift is equal to z the trustor states that her expectation of $\widetilde{a}_2$ is a share equal to b_1. Her underlying belief is then inferred to be $b_1 - z$ and the stated belief b_1 is deemed "admissible" if the underlying belief $b_1 - z$ is in $[0.2, 0.8]$, the interval of transfer shares the trustee can choose from. A stated belief whose underlying belief falls outside this interval is "inadmissible" and indicates a potential confusion

[9] In order to see how this property helps in finding the causal link between b_1 and a_1, consider the simple logic of two-stage least squares regression: the analyst regresses a_1 on z, resulting in a slope coefficient $\beta_{a,z}$, and also regresses b_1 on z, resulting in a coefficient $\beta_{b,z}$. If the only way in which z influences a_1 is through its effect on b_1 (i.e. z and u are orthogonal), it follows that the effect of b_1 on a_1 must be $\frac{\beta_{a,z}}{\beta_{b,z}}$ times as large as the effect of z on b_1. That is, the causal effect of b_1 on a_1 is consistently estimated by $\frac{\beta_{a,z}}{\beta_{b,z}}$.

[10] The twelve tests on the order and laboratory effects all produced p-values above 0.1, with the exception of the test for order effects on the trustee's transfer share in the instrument condition run at York, for which we obtain a p-value of 0.003. Since our data analysis focuses on the trustors, this rejection is not problematic.

Table 2
Means and standard deviations of behavioral variables and shift.

Condition	NI	I (all)	I ("admissible" data)
Trustor's transfer share	0.427 (0.218)	0.435 (0.226)	0.440 (0.227)
Trustee's transfer share	0.306 (0.144)	0.303 (0.150)	0.303 (0.150)
Trustor's belief about adjusted transfer share $\widetilde{a}_2$	–	0.330 (0.185)	0.331 (0.178)
Trustor's belief about transfer share a_2	0.350 (0.132)	–	–
Shift z	– (–)	−0.008 (0.121)	−0.013 (0.119)

on behalf of the trustor subject. We find that only 5 subjects' beliefs (4% of trustors in Condition I) are "inadmissible", a low percentage.[11] For consistency, we exclude these 5 subjects from the analysis, unless mentioned otherwise.

Next, we check whether the instrument has any undesirable effect on the choice variables. The shift's expected value is zero, and thus we check that none of the choice variables exhibits a significant change in means between the treatments with and without the shift. In the game without shift (Condition NI) the transfer shares of the trustors follow a familiar tri-modal pattern that has been observed in many other trust game experiments, with substantial proportions of participants choosing the lowest possible transfer (here, a transfer share of 0.2, chosen by 32.6%), or the midpoint of the action space (0.5 transfer share, chosen by 19.0%) or the highest possible transfer share (0.8, chosen by 14.7%). The remaining observations are dispersed between these three modes. As can be seen in Table 2, the sample mean of the trustor's transfer share is 0.427, with a standard deviation of 0.218. For comparison, in the game with instrument (Condition I) the frequencies of transfer shares that lie on the points of the simple three-point grid $\{0.2, 0.5, 0.8\}$ are 29.9%, 7.7% and 19.7%, and the mean transfer share is 0.435 (std. dev. 0.226).[12]

We conclude that with the single exception of observing fewer transfer shares at level 0.5, the features of the action data under both conditions are very similar. In particular, their means are close to identical and a Mann–Whitney test cannot reject the null hypothesis that the distributions are constant, at any conventional level. The same holds true for the trustees' transfer shares.

Now consider the trustor's belief about the trustee's transfer share. In Condition NI, its mean is 0.350 (std. dev. 0.132), away from its target by less than 5 percentage points. The corresponding numbers for the game with the instrument, Condition I, are close in terms of mean (0.330) but the presence of the instrument induces additional variance in the belief statements—as it should because the shift is random and a rational subject adds the shift to her belief about the opponent. The size of the variance difference is very close to the predicted effect under the assumption that the subjects add the shift to their beliefs.[13]

Fig. 2 provides further evidence that subjects connect underlying beliefs and shifts in an additive way. In the figure, the thick (lower) line represents the target of the belief statements, which is given by the mean behavior of trustees in Condition I plus a trustor's specific value of z. This line's slope is equal to 1 and the set of points on the line corresponds to the set of ex-post optimal belief statements that the trustors would express if their beliefs were rational expectations about a_2. For different underlying beliefs about a_2, optimality requires different levels of belief statements but unvariably with a slope of 1. As the figure shows, a prominent feature of the data is that the large majority of (b_1, z)-pairs indeed lie on straight lines with slope 1: 79.5% of observations in Fig. 2 are (b_1, z)-pairs consistent with the hypothesis that the corresponding participant adds their value of z to a belief about a_2 that lies on the grid $\{0.2, 0.25, 0.3, 0.35, \ldots\}$.[14] The thin (upper) line is the Tobit regression line generated from the depicted data. The fact that the regression line has a slope

[11] Note that expressing an inadmissible belief is a strictly dominated decision, and that their low frequency is actually lower than the frequencies of dominated actions in games, see e.g. Costa-Gomes and Crawford (2006). Additional evidence of a high level of logical consistency of trustors' stated beliefs is provided by the frequencies of observing multiples of 5% in the data, referred to at the end of this subsection.

[12] Note that, in Condition I, if the shift z weakly exceeds 0.14 it is a dominant strategy for the trustor to transfer a share of 0.8 because the amount that she transfers is multiplied by three and she is guaranteed to receive at least 0.34 of this total amount. In the 22 instances with shifts greater than or equal to 0.14, the trustors comply with this prediction 7 times. This is in line with other trust game data where trustors are often found not to trust fully even if it were to pay; see, for example, Bohnet et al. (2005) or Huck et al. (2012).

[13] Under the assumption that the participants in Condition I arrive at their belief statements by simply adding the shift variable z to their (exogenous) belief about a_2, the two variables "belief about a_2" and z are independent and their sum of variances is thus equal to the variance of their sums. Counterfactually assuming that subjects in Condition NI were to observe the same realizations of z and add them to their beliefs, one can analogously construct a variance of "hypothetical beliefs about $\widetilde{a}_2$" in Condition NI, and compare it to the observed variance of beliefs about $\widetilde{a}_2$ in Condition I. The comparison supports the hypothesis that beliefs about the underlying a_2 are constant: the standard deviation of simulated beliefs in Condition NI is 0.181, very close to the standard deviation of stated beliefs in Condition I (0.178).

[14] For comparison, in Condition NI, 90.5% of stated beliefs about the transfer share are multiples of 5%. Stated beliefs about the "adjusted transfer share" $\widetilde{a}_2$ in Condition I are multiples of 5% in only 39.3% of all cases, indicating that many subjects form a well-defined underlying belief on the grid and then add z. For further comparison, Costa-Gomes and Weizsäcker (2008) also find in their 3×3 games that subjects state beliefs that are multiples of 5 percentage points around 90% of the time.

Fig. 2. Trustors' stated beliefs upon observing transfer shifts z. Thick (lower) line: theoretical prediction under rational expectations (slope 1, "target" as described in main text). Thin (upper) line: Tobit regression line (slope 0.841, std. err. 0.116, 95% conf. int. [0.610, 1.072]).

coefficient of 0.841 (std. err. 0.116) that is statistically indistinguishable from 1 is also consistent with the prediction that participants add the instrument to an underlying belief. Finally, a comparison of the second and third columns of Table 2 shows that the exclusion of the inadmissible data does not have much of an effect on the sample statistics.

In sum, the data analyses in this subsection show that introducing the instrument has no undesirable side effects on the distributions of the behavioral variables and that any differences conform to the theoretical predictions of the instrument's effects. We therefore conclude that we can proceed to the IV analysis of data from Condition I and compare its results to the benchmark data from Condition NI.

3.2. Regression analysis: the causal effects of beliefs

In this subsection we present the regression results to assess the causality of beliefs for actions. We write trustor i's transfer share a^*_{1i} as a linear function of her stated belief b_{1i} and a vector of control variables $\mathbf{x}_i$ of self-reported demographic information, socio-economic indicators, cognitive skills, and measures of trust:

$$a^*_{1i} = \beta_0 + \beta_1 b_{1i} + \beta_2 \mathbf{x}_i + u_i \tag{2}$$

Since a^*_{1i} is censored at 0.2 and 0.8 we regard it as a latent variable that underlies the observed value a_{1i},[15]

$$a_{1i} = \begin{cases} 0.2 & \text{if } a^*_{1i} < 0.2 \\ a^*_{1i} & \text{if } 0.2 \le a^*_{1i} \le 0.8 \\ 0.8 & \text{if } a^*_{1i} > 0.8 \end{cases}$$

For the instrumentation in Condition I we also write trustor i's stated belief b_{1i} as a linear function of the transfer shift z_i and control variables.[16] While it would be desirable to do a nonlinear IV analysis of the connection between actions and beliefs, the number of observations only allows us a simple linear specification—with nonlinear models any statistical inference with acceptable levels of power is beyond reach with our data.[17] We first use a two-limit censored Tobit model to estimate the relation between the trustors' transfer shares a_{1i} and their stated beliefs (as in expression (2)) in the NI data, both with and without control variables. This is the analysis that one would carry out in order to establish causality in the absence of endogeneity problems. The results are in Table 3 (standard errors in parentheses; detailed control variables estimates in Table 1A of online Appendix B). The Tobit estimates show a strong correlation of trustors' stated beliefs and their transfer shares, with an average marginal effect of 0.722, see column (1).[18] The corresponding slope coefficient of the linear latent variable a^*_{1i} is 1.317 (std. err. 0.288), but the average marginal effect (calculated via post-regression analysis)

[15] In Condition NI, 31 and 14 observations out of a total of 95 are at the lower and upper limits, respectively. In Condition I, 35 and 23 out of a total of 117 observations are at the lower and upper limits, respectively.

[16] The dependent variable is doubly censored at 0 and 1 but these two belief values appear in less than 5% of the observations.

[17] Under the assumption of self-interested expected-value maximization, the predicted correspondence between beliefs and actions is a nonlinear step function where trustor i chooses the minimal possible share 0.2 if b_{1i} lies below one third and the maximal share 0.8 otherwise. Empirically, 43% and 31% of observations in condition NI and condition I, respectively, conform with this prediction.

[18] In this and subsequent tables, the goodness of fit is measured by $\widehat{R}^2$, denoting the correlation between predicted and the observed values of the dependent variable. The difference in the number of observations between columns (1) and (2) is due to non-response values in the personal questionnaire.

Table 3
Average marginal effects of trustors' beliefs on trustors' on trustors' transfer shares (std. errors in parentheses).

	Transfer share in Condition NI	
	(1) Tobit	(2) Tobit
Belief statement	0.722 (0.126)	0.890 (0.123)
Personal controls	no	yes
# of obs.	95	92
$\widehat{R}^2$	0.214	0.402

Table 4
Average marginal effects of trustors' beliefs on trustors' transfer shares in Condition I (std. errors in parentheses).

	Transfer share in Condition I			
	(1) Tobit	(2) Tobit	(3) IV Tobit	(4) IV Tobit
Belief statement	0.489 (0.102)	0.536 (0.104)	0.519 (0.171)	0.513 (0.190)
Personal controls	no	yes	no	yes
# of obs.	117	116	117	116
$\widehat{R}^2$	0.149	0.254	0.149	0.254

takes into account also the effect of data censoring at transfer shares of 0.2 and 0.8. That is, on average for all observations, including those at the boundary, an increase in the belief of 10 percentage points translates into an increase of 7.2 percentage points in the transfer share. In the regression with controls, the estimated effect is even larger, with a slope of 1.638 (std. err. 0.308) that translates into an average marginal effect of 0.89, see column (2).[19]

A "naive" attribution of these statistical connections to a causal effect would thus suggest that beliefs are a strong driver of trust. The paper's main question is whether this attribution can be corroborated by the IV results. Table 4 has the IV Tobit results from Condition I, showing that the answer is affirmative. As indicated in column (3) of Table 4, the IV average marginal effect from a regression without control variables is estimated at 0.519, with an estimated slope coefficient of 1.004 (std. err. 0.389). This is insignificantly smaller than the non-instrumented estimated in Condition NI, indicating that no strong endogeneity exists.[20] A further indication that there is no strong endogeneity problem is that the size of the IV-estimated effect is very similar to that of the shift itself: In a Tobit regression of a_{1i} on z (see Table A3 in online Appendix B) we estimate an average marginal effect of 0.422 (std. err. 0.168). However, the most important aspect of the IV results of Table 4 is that the IV-estimated effect of beliefs is substantial and significantly different from zero. To the best of our knowledge this is the first direct evidence that the correlation between first-order beliefs and actions in an experimental trust game is indeed causal.

The results also show that within Condition I, there is no discernible difference in the results of Tobit versus IV Tobit. This is another indication that there cannot be a strong omitted-variable problem. One may worry about the observation that the Tobit coefficients differ between Conditions NI and I. The difference is insignificant, however, in a Tobit regression that includes all main and interaction effects of conditions and belief statements.[21]

4. Conclusion

The paper makes two contributions. First, adding to the related literature discussed in the Introduction, it establishes that there is a causal link between first-order beliefs and actions in an investment/trust game. The finding confirms the implicit supposition of such a link in many previous analyses of stated beliefs. The question of causality between beliefs in other people's trustworthiness and actions is potentially relevant for many applied policy issues. Every situation is different, however, and we point out that our "positive" evidence may not generalize to contexts outside the clean laboratory environment.

[19] The estimates in online Appendix B show that age has a statistically significantly negative effect on the transfer share, while the subject's father's level of education, living with a partner, and having a loan as the main source of income have significantly positive coefficients. However, none of these four effects extends to the data from Condition I.

[20] To obtain a statistical test for the comparison, we use the estimated standard deviations of both slope coefficient estimates. The estimates are independent and asymptotically normal. Under the null hypothesis of equal slope coefficients, the standard deviation of the difference between the slope estimates is thus estimated as $\sqrt{0.389^2 + 0.288^2} = 0.484$. The estimated slope difference of $1.317 - 1.004 = 0.313$ is within one estimated standard deviation around zero and has a t-value of $\frac{0.313}{0.484} = 0.647$. Comparing the coefficients from regressions with controls, the analogous standard deviation is $\sqrt{0.408^2 + 0.308^2} = 0.511$ and the slope difference has a t-value of $\frac{1.638-0.959}{0.511} = 1.329$.

[21] To the extent that there is a difference between the two treatments, it could be generated by reciprocity: under Condition NI, trustors may want to be kind to their opponents if they expect them to be kind as well. In Condition I, part of the belief is driven by the computer draw, so a reciprocal agent may respond less to this belief.

Second, the paper further develops and discusses a new method for laboratory experiments—artificially created instruments—that can also be employed to examine other questions. It has always been the hallmark of experimental economics to manipulate directly the explanatory variables of interest, allowing causal insight. Indeed, this is the main reason for why experiments have become so popular. But in some contexts, the explanatory variable of interest is by its very nature an endogenous variable, and thus cannot be fully controlled even by an experimenter. Yet in such contexts, one can at least influence the explanatory variable of interest to some degree, by way of using instrumental variables. Under standard linearity assumptions, this suffices to measure causal links. Non-linear specifications may follow in subsequent research, as may the combination of exogenous randomization with structural-model estimations. Similar procedures to ours may also be applied in studies where the explanatory variable of interest is of a different nature, but is likewise endogenous to the choice process: for example, information about past outcomes, responses to attitudinal questions, happiness reports or even neurological data. To our knowledge, Ham et al. (2005) and Gill and Prowse (2014) are the only previous papers that employ a truly exogenous instrumental variable created in a laboratory—they study the effects of cash balances in auction bidding and, respectively, of past successes in tournament games.[22]

An unusual feature of our study is that we explicitly question the link between expectations and actions—yet traditionally expectations are, at least under subjective expected utility, not viewed as a concept that is separate from actions. We acknowledge that we do not offer an alternative definition of expectations or a general decision-theoretic view on the topic, instead we simply take belief statements as our data. But the statistical establishment of a causal link between expectations and actions is at least pragmatic. Indeed the empirical link may be the only thing that matters for a policy maker who runs a campaign to change expectations in order to accomplish a behavioral change.

Appendix A. An example of naive inference under omitted variables and equilibrium play

In this section we give an example of how the correlation between belief statements and actions can be misleading in the presence of omitted variables. To arrive at a "misleading" effect, we imagine that a researcher observes the full data (choices and belief statements about the opponent's choices) but ignores the possibility of a social norm, or any other unobserved variable, that could drive behavior and belief statements. The players, in contrast, are aware of the full model and play the unique Bayes–Nash Equilibrium (BNE) of the game.

The example builds on a 2×2 mini trust game, where player 1 can either trust ($a_1 = 1$) or not ($a_1 = 0$) and player 2 can reciprocate ($a_2 = 1$) or not ($a_2 = 0$). The players are aware of a social norm that prescribes trust and reciprocation ($a_1 = a_2 = 1$). A random event specifies whether violations of the social norm are sanctioned: in state $\omega = 1$, violations are sanctioned, and we assume that this state arises with probability $\frac{1}{2}$. If $\omega = 1$ occurs, player i's utility is penalized by a term γ_i if she does not comply with the norm but plays $a_i = 0$ instead. The punishment parameter γ_i is known to the player herself but not to her opponent, who only knows the distribution of γ_i to be uniform over $[0, 1]$. If $\omega = 0$, no punishment applies.

A possible justification for such a probabilistic social norm enforcement is that with probability $\frac{1}{2}$ the interaction does not remain anonymous. For example, an outside observer (say, the experimenter) may impose a punishment γ_i on non-cooperative play. Or, the players meet afterwards and may be compelled to reveal their play in the game. In this case, the punishment parameter γ_i would reflect the extent of embarrassment. The payoffs (π_1, π_2) in the two states are as follows.

$\omega = 0$		Player 2	
		$a_2 = 0$	$a_2 = 1$
Player 1	$a_1 = 0$	$0, 0$	$0, 0$
	$a_1 = 1$	$-1, 2$	$1, 1$

$\omega = 1$		Player 2	
		$a_2 = 0$	$a_2 = 1$
Player 1	$a_1 = 0$	$-\gamma_1, -\gamma_2$	$-\gamma_1, 0$
	$a_1 = 1$	$-1, 2 - \gamma_2$	$1, 1$

We assume that the two punishment terms γ_1 and γ_2 are *i.i.d.* uniformly distributed on the interval $[0, 1]$. The worst feasible punishment, $\gamma_i = 1$, makes the non-cooperative action $a_i = 0$ weakly dominated for player i, under state $\omega = 1$. The smallest possible punishment for player 2, $\gamma_2 = 0$, makes player 2's non-cooperative action $a_2 = 0$ weakly dominant (independent of ω). Player 1's optimal action depends on ω, too, but as usual in the trust game it also depends on her belief about a_2—for a large expected return, it pays to trust.

While players do not know the true state ω for sure, they each receive a signal s_i that has precision $\frac{2}{3}$. That is, $\Pr(s_i = 1|\omega = 1) = \Pr(s_i = 0|\omega = 0) = \frac{2}{3}$, for $i = 1, 2$. Their information about ω is therefore correlated: players know that it is more likely than not that the opponent receives the same signal. The probability of the opponent having the same signal is $\frac{5}{9}$ (and the correlation coefficient between the two players' signals is $\frac{1}{9}$).

In this Bayesian game, a player's type is given by her signal s_i and her punishment payoff γ_i. We assume for simplicity that the punishments (γ_1, γ_2) are independent of the signals (s_1, s_2). It is then straightforward to determine the players' optimal choice probabilities: for any signal s_i and any belief about the opponent's strategy, we first ask what values of γ_i

[22] In the literature on field experiments, artificial instruments have been employed to measure the effect of information (Duflo and Saez, 2003) or technology adoption (Devoto et al., 2012) or to avoid selection effects in subsequent experimental interaction of participants (List and Millimet, 2008).

make it optimal for the player to choose the cooperative action $a_i = 1$. The answer yields a cutoff value $\widehat{\gamma}_i(s_i)$, such that for $\gamma_i \geq \widehat{\gamma}_i(s_i)$, the player chooses $a_i = 1$. Each player i employs two such cutoffs, one for each signal realization, $s_i \in \{0, 1\}$. Player i also entertains a belief about the opponent's cooperation: $\Pr(a_j = 1|s_i) = \sum_{\widetilde{s}_j \in \{0,1\}} \Pr(s_j = \widetilde{s}_j|s_i)(1 - \Pr(\gamma_j < \widehat{\gamma}_j(\widetilde{s}_j)))$. This belief determines player i's two cutoffs, and the BNE solution is then found by solving for a set of four cutoffs that form a fixed point. To find the solution, we aggregate over the possible range of punishment parameters and denote the choice probabilities under the players' equilibrium strategies by $r = \Pr(a_1 = 1|s_1 = 0) = 1 - \widehat{\gamma}_1(s_1 = 0)$, $s = \Pr(a_1 = 1|s_1 = 1) = 1 - \widehat{\gamma}_1(s_1 = 1)$, $t = \Pr(a_2 = 1|s_2 = 0) = 1 - \widehat{\gamma}_2(s_2 = 0)$, and $u = \Pr(a_2 = 1|s_2 = 1) = 1 - \widehat{\gamma}_2(s_2 = 1)$. To find e.g. the cutoff value $\widehat{\gamma}_1(s_1 = 1)$ that makes player 1 indifferent upon signal $s_1 = 1$, we solve

$$E\big[\pi_1\big(a_1 = 0\big|s_1 = 1, \widehat{\gamma}_1(s_1 = 1)\big)\big] = E\big[\pi_1\big(a_1 = 1\big|s_1 = 1, \widehat{\gamma}_1(s_1 = 1)\big)\big]$$
$$\frac{2}{3}\big(-\widehat{\gamma}_1(s_1 = 1)\big) = \Pr(a_2 = 0|s_1 = 1) \cdot (-1) + \Pr(a_2 = 1|s_1 = 1) \cdot 1$$

which can be rewritten as:

$$s = \frac{3}{2}\left(\frac{8}{9}t + \frac{10}{9}u\right) - \frac{1}{2}$$

Formulating analogous expressions for r, t and u allows to solve for the unique equilibrium values $\{r = 0, s = \frac{3}{5}, t = \frac{1}{5}, u = \frac{1}{2}\}$. We see that in equilibrium, both players react strongly to their signals as s exceeds r and u exceeds t, both by a considerable margin.[23]

Now consider a naive researcher who wants to infer the causal effect of player 1's beliefs on her actions. We define a naive researcher as one who is not aware that the information structure determines the players' beliefs and actions. Rather, the researcher views the players' beliefs as exogenous and does not require that they are in equilibrium. The researcher collects player-1 data on actions and belief statements about player-2 actions, which we assume are reported truthfully, generated by the full model with social norms. The researcher will therefore observe two different belief statements: first, when player 1 receives the signal $s_1 = 1$, she reports the belief that her opponent cooperates with probability

$$\Pr(a_2 = 1|s_1 = 1) = \frac{5}{9}u + \frac{4}{9}t = \frac{11}{30}.$$

Under this signal realization $s_1 = 1$, we saw above that her actions are cooperative with probability $\frac{3}{5}$. Second, when player 1 receives the signal $s_1 = 0$ she reports that her opponent cooperates with probability

$$\Pr(a_2 = 1|s_1 = 0) = \frac{4}{9}u + \frac{5}{9}t = \frac{1}{3},$$

and her actions under this signal realization are cooperative with probability 0. The data on player 1 that the researcher observes can therefore be summarized in the following table (where the cell entries indicate the relative frequency of the four possible belief-action pairs):

Player 1		Belief statements	
		$bs_1 = \frac{11}{30}$	$bs_1 = \frac{1}{3}$
Actions	$a_1 = 0$	$\frac{2}{10}$	$\frac{1}{2}$
	$a_1 = 1$	$\frac{3}{10}$	0

As the naive researcher ignores the existence of the social norm, he will also wrongly assign causal effects: we assume that he attributes any change in actions exclusively to changes in beliefs. (We also assume that the researcher is not puzzled by the fact that not all actions are best responses to stated beliefs. One could write down a simple error model of what the researcher has in mind, but this would not add much beyond the verbal statement in the sentence before these parentheses.) He therefore believes that if he could intervene and influence players' beliefs, he would also influence players' actions as prescribed by the frequencies in the data matrix. In particular, let us suppose that he thinks he could convince all members of the player-1 population who hold the belief of $\frac{1}{3}$ (i.e. one half of the population) to increase their belief by $\frac{1}{30}$. These player 1s would then hold the same belief as the other half of the population. After such an intervention, the naive

[23] The equilibrium is in (essentially) pure strategies, as a player with a given type has a strict best response, except for the zero-probability event that her realized value γ_i makes her indifferent, i.e. $\gamma_i = \widehat{\gamma}_i(s_i)$.

researcher would expect the actions to change in accordance to the difference between the columns of the above data matrix. He would thus expect the following data after the intervention:

Player 1		Belief statements	
		$bs_1 = \frac{11}{30}$	$bs_1 = \frac{1}{3}$
Actions	$a_1 = 0$	$\frac{2}{5}$	0
	$a_1 = 1$	$\frac{3}{5}$	0

But what would the actual effects be of such an intervention, given the true model? To find the answer, the researcher could use a simple announcement: he could address all player 1s whose belief statement is $\frac{1}{3}$, explaining to them that in one out of 20 times, their opponent would be replaced by a robot that always cooperates.[24] In the above equilibrium, and starting from the belief $\frac{1}{3}$, a player with signal $s_1 = 0$ would indeed arrive at a belief that the opponent cooperates with probability $\frac{11}{30}$, as one can easily check:

$$\begin{aligned}\Pr(\text{opponent cooperates}|s_1 = 0) &= \frac{19}{20}\Pr(a_2 = 1|s_1 = 0) + \frac{1}{20} \\ &= \frac{19}{20}\frac{1}{3} + \frac{1}{20} = \frac{11}{30}\end{aligned}$$

Under the true model, what would be the effect of the announcement on player 1's cooperation rate? What the naive researcher misses is that even under the above announcement, a player 1 with signal $s_1 = 0$ would still assign a low probability to the event that a non-cooperative action would be penalized. She would therefore still find the non-cooperative action $a_1 = 0$ relatively attractive—the omitted variable thus reduces the beneficial effect of the belief shift.

To find the size of the effect, we consider the relevant cutoff $\widehat{\gamma}_1(s_1 = 0)$, after the announcement. The indifference condition is:

$$\begin{aligned}&E\big[\pi_1\big(a_1 = 0\big|s_1 = 0, \widehat{\gamma}_1(s_1 = 0)\big)\big] = E\big[\pi_1\big(a_1 = 1\big|s_1 = 0, \widehat{\gamma}_1(s_1 = 0)\big)\big] \\ &\frac{1}{3}\big(-\widehat{\gamma}_1(s_1 = 0)\big) = \frac{19}{20}\big(\Pr(a_2 = 1|s_1 = 0)1 + \big(1 - \Pr(a_2 = 1|s_1 = 0)\big)(-1)\big) + \frac{1}{20}1 \\ &\frac{1}{3}\big(-\widehat{\gamma}_1(s_1 = 0)\big) = \frac{19}{20}\left(\frac{1}{3} - \frac{2}{3}\right) + \frac{1}{20}1 \\ &\widehat{\gamma}_1(s_1 = 0) = \frac{4}{5}\end{aligned}$$

Thus only a proportion of $\Pr(\gamma_1 \geq \frac{4}{5}) = \frac{1}{5}$ of the players with $s_1 = 0$ would cooperate and the new data matrix after the announcement is

Player 1		Belief statements	
		$bs_1 = \frac{11}{30}$	$bs_1 = \frac{1}{3}$
Actions	$a_1 = 0$	$\frac{3}{5}$	0
	$a_1 = 1$	$\frac{2}{5}$	0

We conclude that by looking at the frequencies instead of measuring the effect, the naive researcher would considerably overestimate the causal link between beliefs and actions. Under the true model, only one fifth of the announcement's recipients would change their actions.

Appendix B. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.geb.2014.10.006.

References

Angrist, J., Pischke, S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.

[24] To be precise, the announcement must be made after the researcher observes the player 1's intended actions and belief statements, but before the game is played. Importantly, for this example, the researcher must not inform player 2 about this intervention, because she would otherwise change her equilibrium behavior. Here in the theoretical example such trickery may be acceptable for the sake of exposition. In our experiments, both players are told about the possibility of intervention, so that no deception is used.

Armantier, O., Treich, N., 2013. Eliciting beliefs: proper scoring rules, incentives, stakes and hedging. Europ. Econ. Rev. 62, 17–40.
Attanasio, O., 2009. Expectations and perceptions in developing countries: their measurement and their use. Amer. Econ. Rev. 99, 87–92 (papers and proceedings).
Bellemare, C., Kröger, S., 2007. On representative social capital. Europ. Econ. Rev. 51, 181–202.
Bellemare, C., Kröger, S., van Soest, A., 2008. Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. Econometrica 76, 815–839.
Bellemare, C., Kröger, S., van Soest, A., 2011a. Preferences, intentions, and expectation violations: a large-scale experiment with a representative subject pool. J. Econ. Behav. Organ. 78, 349–365.
Bellemare, C., Sebald, A., Strobel, M., 2011. Measuring the willingness to pay to avoid guilt: estimation using equilibrium and stated belief models. J. Appl. Econometrics 26, 437–453.
Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games Econ. Behav. 10, 122–142.
Blanco, M., Engelmann, D., Koch, A.K., Normann, H., 2010. Belief elicitation in experiments: is there a hedging problem? Exper. Econ. 13, 412–438.
Bohnet, I., Harmgart, H., Huck, S., Tyran, J.R., 2005. Learning trust. J. Europ. Econ. Assoc. 3, 322–329.
Costa-Gomes, M., Crawford, V., 2006. Cognition and behavior in two-person guessing games: an experimental study. Amer. Econ. Rev. 96, 1737–1768.
Costa-Gomes, M., Weizsäcker, G., 2008. Stated beliefs and play in normal form games. Rev. Econ. Stud. 75, 729–762.
Costa-Gomes, M., Huck, S., Weizsacker, G., 2010. Beliefs and actions in the trust game: creating instrumental variables to estimate the causal effect. ELSE Working Paper 368.
Cox, J., 2004. How to identify trust and reciprocity. Games Econ. Behav. 46, 260–281.
Croson, R., 2000. Thinking like a game theorist: factors affecting the frequency of equilibrium play. J. Econ. Behav. Organ. 41, 299–314.
Devoto, F., Duflo, E., Dupas, P., Pariente, W., Pons, V., 2012. Happiness on tap: piped water adoption in urban Morocco. Amer. Econ. J., Econ. Pol. 4, 68–99.
Duflo, E., Saez, E., 2003. The role of information and social interactions in retirement plan decisions: evidence from a randomized experiment. Quart. J. Econ. 118, 815–842.
Dupas, P., 2011. Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya. Amer. Econ. J., Appl. Econ. 3, 1–34.
Gächter, S., Renner, E., 2010. The effects of (incentivized) belief elicitation in public goods experiments. Exper. Econ. 13, 364–377.
Gill, D., Prowse, V., 2014. Gender differences and dynamics in competition: the role of luck. Quant. Econ. 5, 351–376.
Guiso, L., Sapienza, P., Zingales, L., 2006. Does culture affect economic outcomes? J. Econ. Perspect. 20, 23–48.
Guiso, L., Sapienza, P., Zingales, L., 2009. Cultural biases in economic exchange? Quart. J. Econ. 124, 1095–1131.
Fehr, E., Fischbacher, U., Rosenbladt, B.v., Schupp, J., Wagner, G.G., 2003. A nation-wide laboratory examining trust and trustworthiness by integrating behavioral experiments into representative surveys. IEW Working Paper 141.
Ham, J.C., Kagel, J.H., Lehrer, S.F., 2005. Randomization, endogeneity and laboratory experiments: the role of cash balances in private value auctions. J. Econometrics 125, 175–205.
Huck, S., Lünser, G., Tyran, J.R., 2012. Competition fosters trust. Games Econ. Behav. 76, 195–209.
Huck, S., Weizsäcker, G., 2002. Do players correctly estimate what others do? Evidence of conservatism in beliefs. J. Econ. Behav. Organ. 47, 71–85.
Jensen, R., 2010. The (perceived) returns to education and the demand for schooling. Quart. J. Econ. 125, 515–548.
List, J.A., Millimet, D.L., 2008. The market: catalyst for rationality and filter of irrationality. B.E. J. Econ. Analysis Policy 8 (Article 47).
Manski, C.F., 2004. Measuring expectations. Econometrica 72, 1329–1376.
McKelvey, R.D., Page, T., 1990. Public and private information: an experimental study of information pooling. Econometrica 58, 1321–1339.
Naef, M., Schupp, J., 2009, Measuring trust: experiments and surveys in contrast and combination. Mimeo, Royal Holloway.
Nyarko, Y., Schotter, A., 2002. An experimental study of belief learning using real beliefs. Econometrica 70, 971–1005.
Offerman, T., Sonnemans, J., Schram, A., 1996. Value orientations, expectations and voluntary contributions in public goods. Econ. J. 106, 817–845.
Rutström, E.E., Wilcox, N.T., 2009. Stated beliefs versus inferred beliefs: a methodological inquiry and experimental test. Games Econ. Behav. 67, 616–632.
Sapienza, P., Toldra-Simats, A., Zingales, L., 2013. Understanding trust. Econ. J. 123, 1313–1332.
Servatka, M., Tucker, S., Vadovic, R., 2008, Strategic use of trust. Mimeo, University of Canterbury..
Smith, A., 2013. Estimating the causal effect of beliefs on contributions in repeated public good games. Exper. Econ. 16, 414–425.
Trautmann, S.T., van den Kuilen, G., 2014. Belief elicitation: A horse race among truth serums. Econ. J. http://dx.doi.org/10.1111/ecoj.12160.