Cubitt, Robin and Gächter, Simon and Quercia, Simone (2017) Conditional cooperation and betrayal aversion. Journal of Economic Behavior and Organization, 141 . pp. 110-121. ISSN 0167-2681

Journal of Economic Behavior & Organization 141 (2017) 110–121

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Research paper

Conditional cooperation and betrayal aversion☆

Robin Cubitt[a], Simon Gächter[a,b,c,*], Simone Quercia[d]

[a] University of Nottingham, School of Economics, Sir Clive Granger Building, University Park, Nottingham NG7 2RD, United Kingdom
[b] IZA, Bonn, Germany
[c] CESifo, Munich, Germany
[d] University of Bonn, Institute for Applied Microeconomics, Adenauerallee 24-42, Bonn 53113, Germany

ARTICLE INFO

Article history:
Received 17 August 2015
Received in revised form 16 January 2017
Accepted 24 June 2017
Available online 1 July 2017

JEL classification:
H41
C91
C72
D03

Keywords:
Public goods game
Conditional cooperation
Trust
Betrayal aversion
Exploitation aversion
Free riding
Experiments

ABSTRACT

We investigate whether there is an association between conditional cooperation and betrayal aversion, two phenomena that we conjecture share common psychological characteristics despite having been studied largely separately in the previous literature. We use a public goods game to categorize subjects by type of contribution preference and we measure betrayal aversion for different categories of subject. We report three studies, using two different methods to measure betrayal aversion: a standard elicitation with monetary incentives and a novel scenario-based measure that we argue addresses concerns about the standard measure. We find strong and robust evidence of an association between conditional cooperation and betrayal aversion in the scenario-based measures but not in the standard measure.

1. Introduction

The voluntary provision of public goods is an important economic problem where, notoriously, collective welfare and individual interest diverge. If economic agents are rational and self-interested, public goods will be underprovided, relative to the efficient benchmark (Samuelson, 1954; Hardin, 1968). However, empirical evidence from lab and field studies shows that (i) many people are to some extent willing to provide public goods voluntarily and (ii) those who are willing to do so are mostly "conditional cooperators", that is, they contribute to the public good only if they expect other group members to do so as well (see Ledyard (1995) and Chaudhuri (2011) for overviews).

☆ This work was supported by Economic and Social Research Council [grant number ES/K002201/1], Network for Integrated Behavioural Science and support under ERC-AdG 295707 COOPERATION is gratefully acknowledged. We thank Ben Beranek, Martin Kocher, Felix Kölle, Lucas Molleman, Daniele Nosenzo, Jonathan Schulz, Chris Starmer, Till Weber, Ori Weisel, an editor and two referees for helpful comments.

* Corresponding author at: University of Nottingham, School of Economics, Sir Clive Granger Building, University Park, Nottingham NG7 2RD, United Kingdom.
E-mail addresses: robin.cubitt@nottingham.ac.uk (R. Cubitt), simon.gaechter@nottingham.ac.uk (S. Gächter), simone.quercia@uni-bonn.de (S. Quercia).

http://dx.doi.org/10.1016/j.jebo.2017.06.013

This leaves open what drives people to be conditional cooperators. One psychological motive behind conditional cooperation may be that unconditional cooperation entails the risk of being exploited by free riders; and conditional cooperation protects against this exploitation. This attitude to cooperation suggests that conditional cooperators might be generally reluctant to take *social* risks defined as risky choices where (i) the outcome is caused intentionally by another person and (ii) this person can exploit the risk-taker.

In this paper, we investigate this psychological motive empirically by linking conditional cooperation to another concept that has been closely related to willingness to take social risks of this form, namely *betrayal aversion*. As we discuss in Section 2.1, the idea of betrayal aversion has roots in psychology but was introduced into the economics literature by Bohnet and Zeckhauser (2004) as greater reluctance of people to take *social* risks associated with trusting another person, compared to a benchmark of corresponding *natural* risks (where the outcome is determined by nature, independently of human decisions).[1] A difference in willingness to take these kinds of risk might arise, for example, if people anticipate an emotional cost that they would suffer on top of the material costs, if their trust is betrayed.

We conjecture that, despite being developed largely independently, the concepts of conditional cooperation and betrayal aversion share common psychological characteristics. First, as just indicated each of betrayal aversion and conditional cooperation can be interpreted in terms of reluctance to take social risks as defined above. For example, besides driving betrayal aversion, anticipation of emotional cost of betrayal could be a characteristic of conditional cooperators, as their attitude towards cooperation reveals they are only willing to cooperate conditionally on not being exploited. Second, both conditional cooperation and betrayal aversion are in line with the evidence that subjects care about intentions on top of outcomes (see, e.g., Blount (1995), Falk et al. (2003) and Falk et al. (2008); and for theoretical accounts of such concerns see, e.g., Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2006)). In particular, psychological motives such as reciprocity could explain why people are reluctant to take the risk of intentional exploitation (betrayal aversion, reluctance to contribute if others do not) and why they want to reciprocate by cooperating if others do cooperate (conditional cooperation). While the link between trust and cooperation has been studied before,[2] we are the first, to our knowledge, to investigate how betrayal aversion relates to and inhibits cooperation.

We present the results of three experimental studies that investigate the relation between betrayal aversion and conditional cooperation. Our strategy, common to all three studies, is to measure cooperation preferences using a tool provided by Fischbacher et al. (2001) and to relate them to some measure of betrayal aversion. However, we vary the method used to measure betrayal aversion, starting with a classic approach and then introducing a new method. We discuss the pros and cons of each method in Section 2.2.

Study 1, which we describe in Section 3, uses the techniques of Bohnet and Zeckhauser (2004) to detect betrayal aversion and the tools of Fischbacher et al. (2001) to measure preferences for cooperation and to classify subjects into cooperation "types". We took this approach for our first study in order to replicate the existing classic designs. Studies 2 and 3, which we present in Sections 4 and 5 respectively, introduce a novel form of instrument to measure betrayal aversion that attempts to solve potential problems of the Bohnet and Zeckhauser method, discussed in Section 2.2 and Aimone and Houser (2012).

Our findings provide strong evidence of an association between our second measure of betrayal aversion and cooperation preferences elicited with techniques of Fischbacher et al. (2001). As those techniques are held constant across our analyses in Sections 3–5, we report checks the robustness of our findings in Section 6, by varying and enriching the typology of contribution preferences that Fischbacher et al. (2001) provide. In doing the latter, we address the potential concern that Fischbacher et al. (2001)'s use of "contribution tables" – in which subjects can condition their choices in a public goods game on those of others – may neutralise strategic uncertainty. Since that may attenuate the role of social risk, we also use a further measure of social risk taking in the public goods game. We find evidence that this further measure of social risk taking mediates the relation between conditional cooperation and betrayal aversion in Study 1, but not in Studies 2 and 3. Section 7 summarizes the results and concludes.

2. Conceptualization and measurement of betrayal aversion

2.1. Conceptualization

Like Aimone et al. (2015, p. 2), we conceptualise betrayal aversion as "disutility from the experience, anticipation or observation of non-reciprocated trust", thereby equating betrayal with breach of trust. Such disutility could have several sources, such as emotional cost from experience or anticipation of the act of betrayal, or negative response to presumed intentions of the betrayer. Moreover, as Aimone et al. (2015) point out, the disutility could in principle be inferred from various different types of behavior, including stated disapproval of betrayal (or of betrayers) or avoidance of situations in which betrayal might occur (see, e.g., Koehler and Gershoff (2003)).

[1] Betrayal aversion and, more generally, aversion to social risks have also been investigated by, for example, Koehler and Gershoff (2003), Hong and Bohnet (2007), Bohnet et al. (2008), Gershoff and Koehler (2011), Aimone and Houser (2011), Aimone and Houser (2012), Fetchenhauer and Dunning (2012), Lauharatanahirun et al. (2012), Aimone and Houser (2013), Dreber et al. (2013), Butler and Miller (2014), and Aimone et al. (2015).

[2] See for example Gächter et al. (2004), Thöni et al. (2012) and Balliet and Van Lange (2013) for a meta-analysis.

The approach to detecting betrayal aversion that we take follows that of Bohnet and Zeckhauser (2004) and Bohnet et al. (2008) in inferring the aversion from greater reluctance to take *social* risks than corresponding *natural* risks, where social risks arise from the possibility that another person intentionally exploits the risk taker and natural risks arise from acts of nature. As we explain in more detail below, in Bohnet and Zeckhauser's design, social risk is instantiated by the play of a (human) second mover in a Trust Game and the corresponding natural risk by a move of nature in what is, from the first mover's perspective, an otherwise identical decision problem known as the Risky Dictator Game. In line with this, and as indicated in Section 1, we adopt a conception of social risk that requires *both* intentional human agency *and* the possibility of exploitation.[3] We use such a conception throughout.

In contrast, some authors – such as Lauharatanahirun et al. (2012) – have studied social risks which depend on human agency, but not exploitation. They assess (neural) response to such risks, as compared with corresponding natural risks, by studying the impact of whether or not the risk is framed as resolved by a person or by a roulette wheel. Their implementation of social risk differs from Bohnet and Zeckhauser's games because the person resolving the social risk does not stand to gain from resolving it one way or the other, so the possibility of exploitation is removed. We accept that there are forms of risk that are resolved by social behaviors that are not motivated by desire to exploit. Nevertheless, for our purposes, it is appropriate to confine attention to social risks which do stem from the possibility of exploitation via opportunist acts of others, as our goal is to investigate whether there is a relationship between betrayal aversion and conditional cooperation in relation to public goods. When someone contributes to a public good and finds that others have not, the free riders *have* benefitted in material terms at the expense of the contributor, so exploitation *is* present.

Other authors, such as Koehler and Gershoff (2003), consider aversion to "betrayals" that – at first sight – do not require the betrayer to be human or capable of intention. Koehler and Gershoff (2003) are particularly interested in a form of betrayal that arises when an entity obligated to protect causes the harm that it is supposed to protect from (as, for example, when a security guard steals from her employer or an exploding airbag injures a driver). Though they insist (p. 245) that "Because inanimate objects are incapable of intentionality ., they cannot *really* betray our trust", Koehler and Gershoff (2003) also argue that products such as medicines and safety devices may "*seem* to betray us when they cause the very harms they were designed to guard against" (emphases in original). They term such cases "object betrayals" and report evidence for aversion to them based, for example, on subjects' judgments of appropriate penalties for the suppliers of the malfunctioning products. A natural interpretation – which preserves the principle that only agents capable of intention can *really* be betrayers – is that aversion to "object betrayal" reflects a feeling of being let down by the designer or supplier of the object.[4] However, we do not need to insist on this interpretation or to pursue further the – ultimately philosophical – question of whether inanimate objects can betray. In our studies, we use forms of social risk that arise from intentional acts of conscious agents because these are the kinds of social risk aversion to which might plausibly drive conditional cooperation in public goods problems in human societies.[5]

2.2. Measurement

Consistently with our conceptualisation, we use two different methods to measure betrayal aversion. The first is the method introduced by Bohnet and Zeckhauser (2004). It consists of a between subjects comparison of behavior across two games. In the first game, the Trust Game, a first mover chooses between a safe option and a risky option under which a second mover can intentionally exploit them or reward their trust. The second game, called Risky Dictator Game, has the same structure as the Trust Game except that the outcome of the risky option is determined by nature with probabilities unknown at the time of the first mover's decision.

The design elicits and compares across treatments minimum acceptance probabilities (MAPs), i.e., the minimum probabilities of a favorable outcome from the risky option that the first mover requires to accept that option rather than the safe one. A higher MAP in the Trust Game compared to the Risky Dictator Game was interpreted by Bohnet and Zeckhauser (2004) as evidence for betrayal aversion. To make the elicitation of MAPs incentive-compatible the design uses a version of the Becker-DeGroot-Marschak mechanism (BDM, Becker et al. (1964)): each MAP is compared with a probability p and if the MAP is higher than or equal to p, the first mover is deemed to choose the risky option, while if the MAP is below p they are deemed to choose the safe option. The probability p is equal to the fraction of trustworthy second movers in the Trust Game and to a predetermined and unknown probability in the Risky Dictator Game.

The MAP design has the advantage of being an incentivized behavioral measure, but it presents several potential concerns related both to the BDM mechanism in general and to its specific use in the context of betrayal aversion.

[3] In some (e.g. fictional) contexts, a third element is often present in betrayal, namely breach of an explicit promise. We follow the social science literature in allowing breach of trust *not* to involve broken explicit promises. The complexity of normative expectations that social agents may have of one another precludes seeing them as completely codified in explicit promises.

[4] Besides conforming to the quotations from Koehler and Gershoff (2003) that we have given, this interpretation fits well with the punishments assigned by subjects in their Study 3 being punishments of the supplier of the object that has failed, not of the object itself.

[5] In this discussion, we have conflated intentional human agency and intentional agency, for brevity and for direct applicability to our experiments (in which the only intentional actors are human) and main concerns (human societies). We intend no position on whether trust, betrayal and social risk can feature in non-human societies.

First, from a theoretical standpoint, the BDM mechanism is incentive compatible only under expected-utility theory (see Karni and Safra (1987) and Horowitz (2006) for discussions on the incentive compatibility of the mechanism). Even under the assumption that subjects' 'true' preferences conform to expected utility theory, the mechanism may be empirically unreliable due to misperceptions of the incentive structure (Cason and Plott (2014)).[6]

Second, in the specific setting of betrayal aversion, the mechanism may generate even more confusion than in standard contingent valuation settings as it requires the elicitation of a probability rather than a price, which arguably increases its difficulty (Quercia (2016)).

Third, Aimone and Houser (2012) noticed that if subjects' expectations about the actual level of trustworthiness in the Trust Game differ from their expectations of the unknown probability p in the Risky Dictator Game, differences between MAPs across treatments could arise due to different expectation-based reference points coupled with loss aversion rather than betrayal aversion.

Finally, while the BDM mechanism may be appropriate to elicit consumers' valuations for products, it might be too 'emotionally' cold in the context of a social phenomenon such as betrayal aversion. If, as mentioned in the Introduction, betrayal aversion stems from the anticipation of negative emotional feelings, the highly cognitive nature of the BDM mechanism might obfuscate the salience of these emotional reactions to anticipated betrayal.

Our second measure for betrayal aversion responds to the problems of the MAP method, without departing from our conceptualisation. In particular, we design two hypothetical scenarios ("vignettes") where subjects choose between two taxi companies. In both vignettes, one taxi company is a safe choice as it charges always a fixed fare; while the other company, that charges customers according to the taxi meter, is a risky choice and it can generate a higher or a lower cost (compared to the safe company) with known probabilities. The two vignettes differ just in the source of risk. In one case, the possibility and associated probability of a high cost is due to circumstances outside the driver's control, while in the other case it is due to the intention of the taxi driver to cheat the customer. Our measure of betrayal aversion is the between-subjects difference in the frequency of the safe choice between the two vignettes.

This measure has several advantages. First, it reduces complexity by inferring betrayal aversion from a simple, easy to understand, choice between two commonplace options, rather than from a threshold probability under a complex incentive structure in an unfamiliar decision. Second, we provide subjects with the exact probability of betrayal (bad luck), thus controlling for different expectations across the two vignettes. Third, due to the contextual frame, the cause of betrayal (i.e. intentional exploitation) is made psychologically very salient. Finally, it is worth noticing that betrayal aversion was first elicited experimentally by Koehler and Gershoff (2003) using subjects' responses to simple non-incentivized scenarios where betrayal motives were also psychologically salient. Our second measure goes back to that elicitation principle. Besides reviving an approach from the early betrayal aversion literature in this respect, it also echoes a small but important tradition in economics (see, e.g., Kahneman et al. (1986) for a classic and highly influential study on fairness judgements, using hypothetical scenarios).

3. Study 1–investigating the relation between conditional cooperation and betrayal aversion

3.1. Experimental design and procedures

The design of Study 1 is composed of two parts: Part 1 elicits betrayal aversion replicating the design Bohnet and Zeckhauser (2004); Part 2 measures subjects' attitudes to cooperation following Fischbacher et al. (2001). In each case, we describe game payoffs in each part in terms of "points", for comparability with other studies. (Points were converted to cash at the end of each session.)

3.1.1. Part 1–betrayal aversion

The core of the design of Part 1 is a between-subjects comparison of behavior in two games: the Trust Game and the Risky Dictator Game (henceforth TG and RDG) whose extensive forms are presented in Fig. 1.[7] Subjects are randomly matched in pairs and assigned the roles of first movers and counterpart second movers in TG or, respectively, of first movers and counterpart recipients in RDG. In both games, the first mover chooses between a certain and a risky option. The certain option gives 10 points to him and to his counterpart. The risky option can produce either an unequal outcome of 8 points to the first mover and 22 to his counterpart or an equal split giving 15 points to both.

While in TG the outcome of the risky option is determined by a second mover and the first mover is exposed to the risk of betrayal, in RDG the betrayal and social components are removed by letting the outcome be determined by a random draw with the probabilities of outcomes (15; 15) and (8; 22) being p and $1\text{-}p$, respectively.

[6] For further discussion of the distinction between theoretical and behavioral incentive compatibility, see Bardsley et al. (2010, chapter 6.5).

[7] The original Bohnet and Zeckhauser (2004) design consists of three between-subjects treatments called Trust Game, Risky Dictator Game and Decision Problem. The comparison between choices made in the first and the second treatment measures betrayal aversion, while the comparison between choices made in the second and in the third measures social preferences. As we are mainly interested in measuring betrayal aversion, we implemented only TG and RDG.

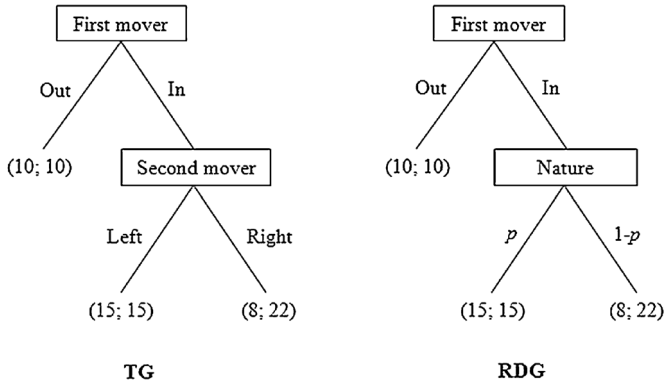

Fig. 1. Extensive forms of the Trust Game (TG) and Risky Dictator Game (RDG).

In our experimental implementation of TG, first and second movers take decisions at the same time; second movers' decisions are elicited using the strategy method (Selten (1967)), i.e., each second mover is asked whether he would choose Left or Right if their first mover chooses In. In RDG, recipients are asked to wait for the decision of first movers and do not take any action.

For first movers in both TG and RDG, we elicit the lowest probability of the outcome (15; 15) at which he would choose In instead of Out. Following Bohnet et al. (2008), we call this value the minimum acceptance probability (MAP). Betrayal aversion is measured as the difference between average MAPs across the two treatments, that is, $BA = \overline{MAP}_{TG} - \overline{MAP}_{RDG}$. If the average MAP in the TG is higher than that in the RDG, first movers are on average betrayal averse.

To elicit MAPs incentive compatibly, we use two different versions of the BDM mechanism: an open-ended version that asks subjects directly their threshold probability and uses the same instructions and procedures as Bohnet et al. (2008)[8] and a choice list version (documented in the web appendix) that uses frequencies instead of probabilities. In both cases, subjects have an incentive to reveal their "true" preference value under expected utility theory (see Bohnet and Zeckhauser (2004) for a discussion of the incentive compatibility of the design). In a companion paper, Quercia (2016) compares these two versions of the BDM mechanism in the context of the elicitation of betrayal aversion. One of the main findings is that the two versions do not produce significantly different levels of betrayal aversion. Given this result, in this paper we pool the data from the two elicitation methods.

3.1.2. *Part 2 – conditional cooperation*

Part 2 uses the design introduced by Fischbacher et al. (2001), in which a variant of the strategy method is employed to elicit subjects' attitudes towards cooperation in a specially designed public goods game. Participants are randomly assigned to groups of four subjects and endowed with 20 tokens, each of which they may either keep for themselves or contribute to a "project". The following payoff function defines the material incentives subjects face:

$$y_i = 20 - x_i + 0.4\sum_{i=1}^{4} x_j$$

where y_i are player i's earnings in points and x_i denotes the contribution of player i to the project. As any token not contributed by a group member increases his/her payoff by 1 point, while any token contributed increases every group member's payoff by 0.4, it is individually optimal to contribute no tokens to the project but socially optimal to contribute every token to it.

Subjects have to make an "unconditional contribution" and a "conditional contribution". In the unconditional contribution, subjects are simply asked how much they contribute to the project. In the conditional contribution task, participants are asked to fill out a contribution table specifying a contribution decision for each possible average contribution (rounded to integers) of the other three subjects in their group. Thus, for each participant their contribution table consists of 21 entries (one for each possible average from 0 to 20). After all participants have completed the unconditional and conditional contribution tasks, we elicit each participant's belief about the average unconditional contribution of the other three members in their group.[9]

[8] This study extends Bohnet and Zeckhauser (2004) including a cross-cultural comparison of the original US sample of Bohnet and Zeckhauser (2004) to five samples from different countries. We draw our instructions which are the same as Bohnet and Zeckhauser (2004) from Bohnet et al. (2008)'s web appendix (http://www.aeaweb.org/aer/data/mar08/20051024_app.pdf). We thank the authors for also making their control questions available.

[9] We follow Fischbacher and Gächter (2010) regarding the mechanism to incentivize beliefs: for each correct belief the participants earned 3 points, for each belief that deviated by 1(2) point from the correct estimate the participants earned 2(1) points.

Table 1
Betrayal aversion according to cooperation types (Std. Dev. in brackets).

	$\overline{MAP}_{TG}$	$\overline{MAP}_{RDG}$	$\overline{MAP}_{TG} - \overline{MAP}_{RDG}$	p-value
All ($n = 273$)	0.52 [0.22]	0.47 [0.20]	0.05	0.031
Conditional cooperators ($n = 163$)	0.53 [0.21]	0.47 [0.20]	0.06	0.109
Free riders ($n = 70$)	0.50 [0.22]	0.43 [0.20]	0.07	0.178
Others ($n = 40$)	0.54 [0.24]	0.50 [0.18]	0.04	0.506

Note: p-values based on Mann-Whitney U (MWU) test.

The incentives are as follows: one group member is randomly selected in each group and their conditional decision is payoff relevant while for the other three group members their unconditional contributions are payoff relevant. The contribution of the randomly selected member is derived calculating the average unconditional contribution of the other three members and finding the corresponding entry in their contribution table. This design has been frequently used to investigate heterogeneity of subjects preferences for cooperation in public goods experiments (see, e.g., Kocher et al. (2008), Herrmann and Thöni (2009), Fischbacher and Gächter (2010), Fischbacher et al. (2012), Martinsson et al. (2013), Frackenpohl et al. (2016)). As we explain below, subjects are divided into types based on their contribution tables.

3.1.3. Procedures

All the sessions were conducted at the University of Nottingham. Participants were recruited via ORSEE (Greiner (2015)). In total, 592 subjects participated in the experiment. Before each part, subjects had to answer a set of control questions to check their understanding of the decision situations. To keep procedures in line with the original studies, Part 1 was run by pen and paper and Part 2 computerized using the software Ztree (Fischbacher (2007)). Subjects were paid the sum of earnings from Part 1 and 2 and all feedbacks about the outcomes of the two parts were given at the end of Part 2 to avoid contamination of Part 2 behavior by Part 1 outcomes. Each session lasted approximately 75 min and the average payment was £ 7.20.

3.2. Experimental results

As a first step, we classify subjects according to their cooperation attitudes elicited in Part 2 of the experiment and we then look at betrayal aversion for each cooperation "type". Based on their public goods contribution tables, we divide subjects into three types: conditional cooperators, free riders, and others. Following Fischbacher et al. (2001), "conditional cooperators" exhibit either a monotonically increasing schedule of contributions or a positive ρ (significant at 1% level) in the Spearman correlation test between the cell-entries of their contribution table and the average contribution levels of the other group members that define the cells. "Free riders" are subjects whose contribution table presents only zeros. All the remaining subjects are classified as "others". In Fischbacher et al. (2001), "triangle contributors" constitute a fourth type. Because there are very few in our sample, we include them in the category "others".

Since we measure betrayal aversion only for the subjects who are assigned the role of first mover in Part 1, the relevant distribution of types for our purposes uses observations only from that subsample ($n = 273$).[10] Consistent with previous literature (see Chaudhuri (2011)), conditional cooperators constitute the majority of our sample (59.7%), followed by free riders (25.6%) and others (14.7%).

Table 1 reports the analysis of betrayal aversion for the entire sample and for each cooperation type. Overall, we observe that betrayal aversion is significant in our sample (MWU-test, $p = 0.031$), but less prominent compared to the original studies by Bohnet and co-authors (the average difference in MAPs between TG and RDG is 0.15 in Bohnet et al. (2008), compared with our figure of 0.05). However, we also notice that other studies using the same design have found smaller sizes of betrayal aversion more in line with our results (0.07 in Hong and Bohnet (2007),[11] 0.08 in Dreber et al. (2013), 0.07 Butler and Miller (2014), and 0.04 in Aimone et al. (2015) using a within-subject version of the design).

Next we look at the relation between betrayal aversion and cooperation preferences from the public goods game. In contrast to our hypothesis, betrayal aversion is not significantly different from zero for any type and, in particular, not for conditional cooperators. A Kruskal-Wallis test on the distribution of MAPs across types cannot reject the null hypothesis of the data being drawn from the same populations in both games ($p = 0.719$ and $p = 0.327$ for TG and RDG, respectively).

3.3. Discussion

The results of Study 1 suggest that betrayal aversion is not specifically a characteristic of conditional cooperators, as opposed to other types. Yet, as discussed in Section 2, one potential limitation of this study is its vulnerability to concerns

[10] This number also excludes double-switchers in the choice list table. For these subjects we are not able to infer a MAP. The distribution of types of these 273 subjects is not significantly different from the distribution of types for subjects who were in the role of second movers (recipients) in Part 1 ($\chi^2(2) = 3.30$, $p = 0.192$). Thus, we can exclude any spillover effect of the role subjects were playing in Part 1 on Part 2 behavior.

[11] This statistic is not reported in the paper. We thank the authors for providing their data.

Natural risk vignette
For personal reasons, you have to travel to a big city. From the airport you can choose between two taxi companies to reach your final destination for which you don't know the exact route. Company A charges you a fixed price of $12. Company B charges you according to the taxi meter. If the weather is fine, it costs you $8. However, 1 out of 5 times, due to bad weather conditions the ride takes longer and the fare is then $16.

Social risk vignette
For personal reasons, you have to travel to a big city. From the airport you can choose between two taxi companies to reach your final destination for which you don't know the exact route. Company A charges you a fixed price of $12. Company B charges you according to the taxi meter. If the driver takes the direct route, it costs you $8. However, 1 out of 5 drivers take detours to make more money out of you and the fare is then $16.

about using the BDM mechanism to elicit betrayal aversion, such as that (i) the BDM might not be transparent to subjects and hence generate noisy valuations, (ii) the obtained measure of betrayal aversion might be confounded (e.g. with loss aversion) and (iii) the salience of the psychological mechanism driving betrayal aversion might be rather low as the BDM mechanism is arguably very cognitively demanding. In the next section, we introduce Study 2 where we address these concerns by investigating our research question with our second measure of betrayal aversion.

4. Study 2–a novel measure of betrayal aversion

For Study 2, we developed vignettes where we ask subjects to choose between a safe option and a risky option either in a social risk situation with betrayal possibility or in a natural risk situation. The between-subjects comparison of hypothetical willingness to take risks in these two scenarios constitutes our second measure of betrayal aversion.

4.1. Experimental design and procedures

Study 2 was conducted online using Amazon Mechanical Turk (see Horton et al. (2011)). The experiment was composed of two one-shot experiments. The first was a strategy method public goods game experiment identical to the one used in Study 1. The second game was a direct response one-shot public goods game where we also elicited incentivized beliefs about the average contribution of the other three group members. After subjects made their decisions in these two games, they were presented with one of the two following vignettes (bold font and titles were not presented to subjects):

After reading, subjects were asked to choose one of the two companies. Notice that a risk neutral, cost minimizing, agent would always choose Company B, which has an expected cost of $9.60. However, depending on their degree of risk aversion, some decision makers may choose Company A in the natural risk vignette. We expect that the possibility of betrayal would make our subjects more likely to choose Company A in the social risk vignette than they are in the natural risk vignette. We will interpret such a finding as evidence of betrayal aversion. In the spirit of Kahneman and Tversky (1979, p. 265), our use of responses to hypothetical scenarios in this way rests on the premise that most subjects will (i) understand the decision problems described; (ii) know what they would choose in them and (iii) have no reason to mislead us about these choices. We think it likely that all three conditions are met in our study.

In total, we had 359 subjects who were paid $2 for participating plus an additional bonus that depended on their earnings from the two public goods games. The average bonus earnings were $ 2.56 and the experiment lasted on average 9 min.

4.2. Experimental results

The results from the classification of cooperation types show a very high proportion of conditional cooperators compared to the laboratory sample of Study 1. In particular, we find 79.7% conditional cooperators, 8.6% free riders and 11.7% others. This is consistent with previous lab experiments on US subject pools (see e.g. Kocher et al. (2008)[12]).

The results from the vignettes reported in Table 2 indicate that, overall, participants are significantly betrayal averse, as shown by the percentage points difference in subjects choosing Company A (the safe option) between the two scenarios: the percentage increases from 45% in the natural risk vignette to 63.7% in the social risk one ($\chi^2(1) = 12.63$, $p < 0.001$). In Table 2, we also report the percentage of subjects in each type sample who choose Company A, separately for natural and social risk vignettes. Statistical comparison reveals significant betrayal aversion only for conditional cooperators and others, but not for free riders.

[12] In particular, they find 80.6% conditional cooperators, 8.3% free riders and 11.1% others. This distribution is not statistically different from ours ($\chi^2(2) = 0.016$, $p = 0.992$), which is reassuring for our methodology of using online experiments for this study. We thank the authors for supplying their data.

Table 2
Percentage choosing the safe option (Company A) for each cooperation type.

	Social Risk	Natural Risk	Percentage points difference	*p*-value
All ($n = 359$)	64%	45%	19	<0.001
Conditional cooperators ($n = 286$)	63%	47%	16	0.007
Free riders ($n = 31$)	50%	24%	26	0.125
Others ($n = 42$)	80%	50%	30	0.043

Note: *p*-values based on χ^2 test.

Table 3
Percentage of subjects choosing Company A – Study 3.

	Social Risk	Natural Risk	Percentage points difference	*p*-value
All ($n = 600$)	57%	47%	10	0.014
Conditional cooperators ($n = 458$)	60%	46%	14	0.004
Free riders ($n = 84$)	31%	36%	−5	0.643
Others ($n = 58$)	72%	77%	−5	0.662

Note: *p*-values based on χ^2 test.

4.3. Discussion

These results reveal that, in Study 2, subjects are overall betrayal averse and conditional cooperators are betrayal averse, as hypothesized. However, Study 2 has some limitations. First of all, the results are not conclusive for free riders and others due to the small number of observations in these categories. Second, the vignette task was always conducted after the public goods game with the potential danger of order or congruence effects. In Study 3, which we present in the next section, we tackle these two problems and also replicate Study 2.

5. Study 3–Implementing our novel measure of betrayal aversion in a large sample

Study 3 uses the same design and vignettes of Study 2, but we present the tasks (public goods games and vignettes) in two orders manipulated between subjects. Thus, we have a 2×2 design that varies order and type of scenario (natural vs. social risk). In the first order, subjects made decisions for the vignettes first and then in the public goods games; the second order was like our Study 2 where subjects made choices for the public goods game first and then for the vignettes. The experiment was conducted online on Amazon Mechanical Turk with a large sample of 600 participants to allow us to make inferences on free-riders and others. Incentives were the same as in Study 2.

5.1. Experimental results

First, we investigate order effects keeping fixed the type of vignette. In the natural risk vignette, we find 45% of subjects choosing the safe option (Company A) when the vignette is presented before the public goods game and 49% when it is presented afterwards. This difference is not significant at conventional levels ($n = 301$, $\chi^2(1) = 0.407$, $p = 0.523$). We obtain a similar result in the social risk vignette as 57% of subjects choose the safe option both when the vignette is presented before and after the public goods game ($n = 299$, $\chi^2(1) = 0.007$, $p = 0.933$). Given these results, we pool the two orders in the data analysis below.

Next, we look at betrayal aversion in the whole sample. We find 57% subjects choosing the safe option in the social risk vignette and 47% in the natural risk vignette ($n = 600$, $\chi^2(1) = 6.029$, $p = 0.014$), indicating subjects are more averse to risks in the social than in the natural risk vignette replicating the results of Study 2.[13] With respect to classification of types in the public goods game we find 76% conditional cooperators, 14% free riders and 10% others. Thus, Study 3 has a substantially larger number of subjects of each type than Study 2.

In Table 3 we report betrayal aversion for the entire sample and for each type in the public goods game. Choice proportions of conditional cooperators are very similar to Study 2 (compare Table 2). Conditional cooperators are significantly betrayal averse. Free riders and others are not. This finding confirms the link between betrayal aversion and conditional cooperation found in Study 2.

5.2. Discussion

The results from Study 3 replicate the results of Study 2 for overall betrayal aversion and for betrayal aversion of conditional cooperators. Notwithstanding the larger sample size, they do not provide evidence of betrayal aversion for either

[13] Each percentage is not significantly different from the corresponding one found in Study 2 (social risk: $\chi^2(1) = 1.963$, $p = 0.161$; natural risk: $\chi^2(1) = 0.215$, $p = 0.643$).

free riders or others. Overall the results from Studies 2 and 3 confirm the hypothesised association between conditional cooperation and betrayal aversion.

6. Alternative categorizations of cooperation types

One feature of the analyses reported up to now is that they all use the measurement techniques and typology of Fischbacher et al. (2001). In this section, we investigate the robustness of this approach by (i) allowing different classification of individuals' attitudes to cooperation and (ii) enriching the classification with an index of willingness to take social risk in the public goods game.

The first exercise is intended to check the robustness of results to variations in the way we classify cooperation types in the public goods game. In particular, we define free riders using weaker criteria. In the original Fischbacher et al. (2001) classification, a subject is classified as free rider only if they fill their contribution table with zeros. We now consider two alternative classifications that allow "free riders" to be mildly or occasionally cooperative. In one, we classify subjects as free riders if their *maximum* entry in the contribution table is less than 3 tokens and, in the other, if their *average* entry in the contribution table is less than 3 tokens.[14] The classification criteria for conditional cooperators are unchanged and if a subject fits with both conditional cooperators and free rider criteria, we classify them as a free rider. Subjects are classified as 'others' if they are categorized neither as a free rider nor as a conditional cooperator.

We conduct the same analyses as are reported in Sections 3–5 and report the results in Tables A1–A6 in the Appendix A. All main results reported in previous sections are replicated according to the new classification criteria. As before, we do not find significant betrayal aversion for any of the types in Study 1. In Study 2, conditional cooperators are significantly betrayal averse no matter how we change the definition of free riders. Free riders are significantly betrayal averse according to classification (i) above but not according to (ii). Others are mildly betrayal averse in Study 2. As stressed in Section 4 however, the results for free riders and others in Study 2 must be taken carefully because of the low number of observations. In Study 3, conditional cooperators are significantly betrayal averse under all classification criteria while free riders and others are not (see Tables A1–A5 in the Appendix A). We conclude that our results are robust to changes in the classification of free rider types considered.

Next, we ask whether considering only cooperation attitudes might miss some elements of the link between conditional cooperation and betrayal aversion. In particular, it is arguable that cooperation attitudes elicited with the Fischbacher et al. (2001) conditional contribution tasks do not reveal willingness to take social risks as the strategy method removes strategic uncertainty from the public goods game by, in effect, structuring the game as sequential. This could be a concern in the context of establishing a link between conditional cooperation and betrayal aversion, when the latter interpreted as a behavioral trait that inhibits the taking of social risks.

As a direct measure of social risk taking in the public goods game, we now use the unconditional contributions elicited in the Fischbacher et al. (2001) design and combine them with the standard classification of subjects obtained from the conditional decisions. (For clarity, we stress that we will still refer to subjects as conditional cooperators, if they are so classified on the basis of their contribution tables.) Unconditional contributions reflect how much subjects expose themselves to the risk of exploitation by groupmates who contribute less than they do. We expect stronger betrayal aversion among conditional cooperators who contribute little unconditionally than among those who contribute more unconditionally, as low unconditional contributions protect against exploitation, while high unconditional contributions leave the subject exposed to the risk of being exploited.

We investigate this conjecture using OLS regressions. For Study 1, we regress the variable MAP (minimum acceptance probability) on a treatment dummy SR (social risk; =1 if the observation is collected in TG), on the variable Unconditional Contribution and on the interaction term between SR and Unconditional Contribution. For each of Studies 2 and 3, we regress the variable SAFE (=1 if the subjects chose Company A, i.e., the safe company in our vignettes) on a treatment dummy SR (=1 if subject was exposed to the social risk vignette), on the variable Unconditional Contribution and on the interaction term between SR and Unconditional Contribution. For each study, we report the results for conditional cooperators in Table 4.[15]

We are mainly interested in the treatment variable (SR) that reveals betrayal aversion for individuals contributing 0 tokens unconditionally and the interaction term (SR × Unconditional Contribution) indicating how betrayal aversion changes as the unconditional contribution increases. In all three studies, we find the treatment dummy to be positive and significant (although at the 10% level in Studies 2 and 3). This indicates that conditional cooperators who contribute 0 unconditionally are betrayal averse in all studies. If betrayal aversion decreases as social risk taking increases, as hypothesized, we should expect the interaction term to be negative and significant. This is confirmed in Study 1 but, despite the sign being in the

[14] These modifications are sensible as in the original Fischbacher et al. (2001)'s classification, some subjects who are arguably closer to a free riding behavior are classified as conditional cooperators as for example subjects who have only one monotonic increase by one token and no decrease in their contribution schedule. We thank the editor for suggesting these alternative classification criteria.

[15] We do not use Probit or Logit models in our Studies 2 and 3 despite having a binary dependent variable as we are mainly interested in the interaction term in the regression. As noted by Ai and Norton (2003), the coefficient of the interaction term using Probit and Logit models would be not interpretable using a standard software package; hence, we prefer using linear probability models (see also Angrist and Pischke (2008)).

Table 4
The relation between conditional cooperation, unconditional contributions and betrayal aversion.

Dependent variable	Study 1 MAP	Study 2 SAFE	Study 3 SAFE
SR	0.140** (0.055)	0.203* (0.108)	0.140* (0.077)
Unconditional Contribution	0.005 (0.003)	-0.004 (0.006)	0.005 (0.004)
SR × Unconditional Contribution	-0.009* (0.005)	-0.004 (0.008)	-0.0002 (0.006)
Constant	0.427*** (0.035)	0.516*** (0.082)	0.412*** (0.056)
N	163	286	458

Notes: OLS regressions; dependent variables: MAP (minimum acceptance probability) in Study 1 and SAFE dummy (=1 if Company A is chosen) for Study 2 and 3.* significant at 10%, ** significant at 5%, *** significant at 1%. Robust standard errors in parentheses.

Table A1
Robustness check 1–Study 1, free riders defined as *maximum* conditional contribution less than 3 tokens.

	$\overline{MAP}_{TG}$	$\overline{MAP}_{RDG}$	$\overline{MAP}_{TG} - \overline{MAP}_{RDG}$	p-value
Conditional cooperators ($n = 156$)	0.53 [0.21]	0.48 [0.20]	0.05	0.161
Free riders ($n = 80$)	0.49 [0.22]	0.43 [0.21]	0.06	0.197
Others ($n = 37$)	0.57 [0.21]	0.50 [0.19]	0.07	0.330

Note: average MAPs [Std. Dev.]; p-values based on Mann-Whitney U (MWU) test.

right direction in Studies 2 and 3, the interaction term is not significant in those cases.[16] Hence, we conclude that our second measure of betrayal aversion does not seem to be related to the willingness to take social risk as reflected by unconditional contributions.

7. Summary and conclusion

In this paper, we have reported the results of three studies that investigate whether there is a link between betrayal aversion and conditional cooperation, two concepts studied up to now in distinct strands of the experimental literature. We have employed two different methods to measure betrayal aversion. In Study 1, we relied on a tool introduced in the economics literature by Bohnet and Zeckhauser (2004) that measures betrayal aversions using a version of the Becker-DeGroot-Marschak mechanism. In Studies 2 and 3, we introduced a novel method based on vignettes that simplifies substantially the measurement and elicitation of betrayal aversion.

While we do not find support for an association between conditional cooperation and betrayal aversion in Study 1, we do find it in Study 2 and in Study 3. In Section 6, we have confirmed the robustness of our results to alternative classifications of subjects' dispositions in public goods games. As discussed in Section 2, we think that our use of vignettes has important advantages over the Becker-DeGroot-Marschak mechanism. To this extent, our findings provide strong evidence of an association between conditional cooperation and betrayal aversion. Of course, they do not indicate a specific direction of causality. Nevertheless, we conjecture that, rather than there being a unidirectional causal chain from one phenomenon to the other, each of conditional cooperation and betrayal aversion reflects a more basic aversion to intentional exploitation.

Appendix A.

In this Appendix we report several robustness checks as described in Section 6. Tables A1 and A2 should be compared to Table 1; Tables A3 and A4 should be compared to Table 2; and Tables A5 and A6 to Table 3. In each case, as indicated by its header, the table below uses a different typology of subjects than the one used in the main text.

Table A7 should be compared to Table 4. It uses subjects' beliefs about the unconditional contributions of others in place of their own unconditional contribution, as an alternative indicator of social risk taking in the public goods game.

[16] An alternative approach is to use beliefs about the average contribution of the three other group members as a measure of social risk taking. For conditional cooperators, consistent with their attitude, unconditional contributions and beliefs are highly correlated (correlation coefficients are 0.77, 0.55, 0.65 in Studies 1, 2, and 3, respectively; all p-values <0.001). The results using beliefs instead of unconditional contributions are similar to the ones reported here and reported in Table A7 in the Appendix A.

Table A2
Robustness check 2–Study 1, free riders defined as *average* conditional contribution less than 3 tokens.

	$\overline{MAP}_{TG}$	$\overline{MAP}_{RDG}$	$\overline{MAP}_{TG} - \overline{MAP}_{RDG}$	p-value
Conditional cooperators ($n = 135$)	0.52 [0.20]	0.48 [0.20]	0.04	0.346
Free riders ($n = 114$)	0.51 [0.23]	0.45 [0.20]	0.06	0.113
Others ($n = 24$)	0.59 [0.23]	0.48 [0.17]	0.11	0.283

Note: average MAPs [Std. Dev.]; p-values based on Mann-Whitney U (MWU) test.

Table A3
Robustness check 1–Study 2, free riders defined as *maximum* conditional contribution less than 3 tokens.

	Social Risk	Natural Risk	Percentage points difference	p-value
Conditional cooperators ($n = 284$)	62%	47%	15	0.009
Free riders ($n = 34$)	56%	22%	34	0.042
Others ($n = 41$)	80%	52%	28	0.062

Note: percentage of subjects choosing Company A; p-values based on χ^2 test.

Table A4
Robustness check 2–Study 2, free riders defined as *average* conditional contribution less than 3 tokens.

	Social Risk	Natural Risk	Percentage points difference	p-value
Conditional cooperators ($n = 270$)	63%	45%	18	0.004
Free riders ($n = 50$)	55%	39%	16	0.283
Others ($n = 39$)	80%	53%	27	0.070

Note: percentage of subjects choosing Company A; p-values based on χ^2 test.

Table A5
Robustness check 1–Study 3, free riders defined as *maximum* conditional contribution less than 3 tokens.

	Social Risk	Natural Risk	Percentage points difference	p-value
Conditional cooperators ($n = 455$)	60%	46%	14	0.002
Free riders ($n = 88$)	29%	36%	−7	0.492
Others ($n = 57$)	72%	80%	−8	0.479

Note: percentage of subjects choosing Company A; p-values based on χ^2 test.

Table A6
Robustness check 2–Study 3, free riders defined as *average* conditional contribution less than 3 tokens.

	Social Risk	Natural Risk	Percentage points difference	p-value
Conditional cooperators ($n = 420$)	61%	45%	16	0.001
Free riders ($n = 134$)	41%	43%	−2	0.794
Others ($n = 46$)	67%	81%	−14	0.242

Note: percentage of subjects choosing Company A; p-values based on χ^2 test.

Table A7
The relation between conditional cooperation, beliefs and betrayal aversion.

Dependent variable	Study 1 MAP	Study 2 SAFE	Study 3 SAFE
SR	0.151** (0.062)	0.263** (0.131)	0.131 (0.101)
Beliefs	0.006 (0.004)	−0.004 (0.009)	0.003 (0.006)
SR × Beliefs	−0.010* (0.005)	−0.011 (0.012)	0.0003 (0.009)
Constant	0.417*** (0.040)	0.507*** (0.102)	0.429*** (0.071)
N	163	286	458

Notes: OLS regressions; dependent variables: MAP (minimum acceptance probability) in Study 1 and safe choice dummy (=1 if Company A is chosen) for Study 2 and 3.* significant at 10%, ** significant at 5%, *** significant at 1%. Robust standard errors in parentheses.

Appendix B. Supplementary data

Supplementary data experimental instructions, software, data and analysis files associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jebo.2017.06.013.

References

Ai, C., Norton, E.C., 2003. Interaction terms in logit and probit models. Econ. Lett. 80 (1), 123–129.
Aimone, J.A., Houser, D., 2011. Beneficial betrayal aversion. PLoS One 6 (3), e17725.
Aimone, J.A., Houser, D., 2012. What you don't know won't hurt you: a laboratory analysis of betrayal aversion. Exp. Econ. 15 (4), 571–588.
Aimone, J.A., Houser, D., 2013. Harnessing the benefits of betrayal aversion. J. Econ. Behav. Org. 89, 1–8.
Aimone, J., Ball, S., King-Casas, B., 2015. The betrayal aversion elicitation task: an individual level betrayal aversion measure. PLoS One 10 (9), e0137491.
Angrist, J.D., Pischke, J.-S., 2008. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.
Balliet, D., Van Lange, P.A.M., 2013. Trust, conflict, and cooperation: a meta-analysis. Psychol. Bull. 139 (5), 1090–1112.
Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., Sugden, R., 2010. Experimental Economics: Rethinking the Rules. Princeton University Press, Princeton.
Becker, G.M., DeGroot, M.H., Marschak, J., 1964. Measuring utility by a single-response sequential method. Behav. Sci. 9 (3), 226–232.
Blount, S., 1995. When social outcomes aren't fair—the effect of causal attributions on preferences. Organ. Behav. Hum. Decis. Process. 63 (2), 131–144.
Bohnet, I., Zeckhauser, R., 2004. Trust, risk and betrayal. J. Econ. Behav. Org. 55 (4), 467–484.
Bohnet, I., Greig, F., Herrmann, B., Zeckhauser, R., 2008. Betrayal aversion. Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. Am. Econ. Rev. 98 (1), 294–310.
Butler, J.V., Miller, J.B., 2014. Social Risk: The Role of Warmth and Competence. Working Paper Series IGIER. Bocconi University, n. 522.
Cason, T.N., Plott, C.R., 2014. Misconceptions and game form recognition: challenges to theories of revealed preference and framing. J. Polit. Econ. 122 (6), 1235–1270.
Chaudhuri, A., 2011. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. Exp. Econ. 14 (1), 47–83.
Dreber, A., Rand, D.G., Wernerfelt, N., Worrell, P.R., Zeckhauser, R.J., 2013. The Decisions of Entrepreneurs and Their Agents: Revealed Levels of Risk Aversion and Betrayal Aversion. Harvard Kennedy School Research Working Paper Series – RWP 13-016.
Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. Games Econ. Behav. 47 (2), 268–298.
Falk, A., Fischbacher, U., 2006. A theory of reciprocity. Games Econ. Behav. 54 (2), 293–315.
Falk, A., Fehr, E., Fischbacher, U., 2003. On the nature of fair behavior. Econ. Inq. 41 (1), 20–26.
Falk, A., Fehr, E., Fischbacher, U., 2008. Testing theories of fairness-Intentions matter. Games Econ. Behav. 62 (1), 287–303.
Fetchenhauer, D., Dunning, D., 2012. Betrayal aversion versus principled trustfulness—how to explain risk avoidance and risky choices in trust games. J. Econ. Behav. Org. 81 (2), 534–541.
Fischbacher, U., Gächter, S., 2010. Social preferences, beliefs, and the dynamics of free riding in public good experiments. Am. Econ. Rev. 100 (1), 541–556.
Fischbacher, U., Gächter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. Econ. Lett. 71 (3), 397–404.
Fischbacher, U., Gächter, S., Quercia, S., 2012. The behavioral validity of the strategy method in public good experiments. J. Econ. Psychol. 33 (4), 897–913.
Fischbacher, U., 2007. z-Tree: Zurich toolbox for readymade economic experiments. Exp. Econ. 10 (2), 171–178.
Frackenpohl, G., Hillenbrand, A., Kube, S., 2016. Leadership effectiveness and institutional frames. Exp. Econ. 19 (4), 842–863.
Gächter, S., Herrmann, B., Thöni, C., 2004. Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence. J. Econ. Behav. Org. 55 (4), 505–531.
Gershoff, A.D., Koehler, J.J., 2011. Safety first? The role of emotion in safety product betrayal aversion. J. Consum. Res. 38 (1), 140–150.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1 (1), 114–125.
Hardin, G., 1968. The tragedy of the commons. Science 162 (3859), 1243–1248.
Herrmann, B., Thöni, C., 2009. Measuring conditional cooperation: a replication study in Russia. Exp. Econ. 12 (1), 87–92.
Hong, K., Bohnet, I., 2007. Status and distrust: the relevance of inequality and betrayal aversion. J. Econ. Psychol. 28 (2), 197–213.
Horowitz, J.K., 2006. The Becker-DeGroot-Marschak mechanism is not necessarily incentive compatible, even for non-random goods. Econ. Lett. 93 (1), 6–11.
Horton, J.J., Rand, D.G., Zeckhauser, R.J., 2011. The online laboratory: conducting experiments in a real labor market. Exp. Econ. 14 (3), 399–425.
Kahneman, D., Tversky, A., 1979. Prospect theory—analysis of decision under risk. Econometrica 47 (2), 263–291.
Kahneman, D., Knetsch, J.L., Thaler, R., 1986. Fairness as a constraint on profit seeking: entitlements in the market. Am. Econ. Rev. 76 (4), 728–741.
Karni, E., Safra, Z., 1987. Preference reversal and the observability of preferences by experimental methods. Econom.: J. Econom. Soc., 675–685.
Kocher, M.G., Cherry, T., Kroll, S., Netzer, R.J., Sutter, M., 2008. Conditional cooperation on three continents. Econ. Lett. 101 (3), 175–178.
Koehler, J.J., Gershoff, A.D., 2003. Betrayal aversion: when agents of protection become agents of harm. Organ. Behav. Hum. Decis. Process. 90 (2), 244–261.
Lauharatanahirun, N., Christopoulos, G.I., King-Casas, B., 2012. Neural computations underlying social risk sensitivity. Front. Hum. Neurosci. 6, 213.
Ledyard, J.O., 1995. Public goods: a survey of experimental research. In: Roth, A.E., Kagel, J.H. (Eds.), The Handbook of Experimental Economics. Princeton University Press, Princeton, pp. 111–181.
Martinsson, P., Pham-Khanh, N., Villegas-Palacio, C., 2013. Conditional cooperation and disclosure in developing countries. J. Econ. Psychol. 34 (0), 148–155.
Quercia, S., 2016. Eliciting and measuring betrayal aversion using the BDM mechanism. J. Econ. Sci. Assoc. 2 (1), 48–59.
Samuelson, P.A., 1954. The pure theory of public expenditure. Rev. Econ. Stat. 36 (4), 387–389.
Selten, R., 1967. Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. In: Sauermann, H. (Ed.), Beiträge Zur Experimentellen Wirtschaftsforschung, vol. I. Mohr Tübingen, pp. 136–168.
Thöni, C., Tyran, J.-R., Wengström, E., 2012. Microfoundations of social capital. J. Publ. Econ. 96 (7–8), 635–643.