

Wolff, Irenaeus

**Conference Paper**

## Predicting Voluntary Contributions by 'Revealed-Preference Nash-Equilibrium'

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2022: Big Data in Economics

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Wolff, Irenaeus (2022) : Predicting Voluntary Contributions by 'Revealed-Preference Nash-Equilibrium', Beiträge zur Jahrestagung des Vereins für Socialpolitik 2022: Big Data in Economics, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/264072>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Predicting Voluntary Contributions by 'Revealed-Preference Nash-Equilibrium'<sup>§</sup>

Irenaeus Wolff

*Thurgau Institute of Economics (TWI) / University of Konstanz  
Hafenstrasse 6, 8280 Kreuzlingen, Switzerland  
wolff@twi-kreuzlingen.ch*

---

**Abstract:** One-shot public-good situations are prominent in the public debate, and a prime example for behaviour diverging from the standard Nash-equilibrium. But does behaviour diverge from equilibrium because the equilibrium fails to account for social preferences or—as recent research suggests—because a Nash-equilibrium cannot possibly predict one-shot public-good behaviour? I show that a 'revealed-preference Nash-equilibrium' (RPNE) out-of-sample predicts one-shot public-good behaviour, outperforming other social-preference models. The RPNE is the set of 'mutual conditional contributions', interpreting elicited conditional contributions as best-responses.

Individual-level analyses confirm the results and allow for studying equilibrium selection. The modal choice corresponds to the Pareto-dominant equilibrium. However, many participants use other criteria. Given the predictive positive-contributions RPNEs, the heterogeneity in selection criteria may turn out to be the origin of the literature's low late-round contribution levels. Conversely, many real-life public-good situations may be solvable at least partially if the players can coordinate on an equilibrium-selection criterion beforehand.

*Keywords:* Social dilemma, public good, conditional cooperation, Nash-equilibrium, best-response, social preferences, preference stability, knowledge of preferences.

*JEL:* C72, C92, D83, H41

---

<sup>§</sup>I am grateful to Jordi Brandts, Gary Charness, Holly Dykstra, Sebastian Fehrler, Urs Fischbacher, Simon Gächter, Botond Köszegi, Rachel Kranton, Wieland Müller, Heinrich Nax, Louis Putterman, Dirk Sliwka, Jean-Robert Tyran, Marie Claire Villeval, Alexander K. Wagner, Roberto Weber, Ro'i Zultan and the research group at the Thurgau Institute of Economics (TWI) for fruitful discussions on the project and/or comments on earlier drafts. I further thank seminar audiences at the Universities of Vienna, Innsbruck, and Nottingham, and the participants of several conferences. Funding by SNF grant 100018\_152788 is gratefully acknowledged.

## 1 Introduction

One-shot public-good situations are extremely prominent in both economic textbooks and popular conceptualisations of some of the most pressing problems humanity is facing (*e.g.*, climate change). At the same time, one-shot public-good experiments are a prime example for a situation in which people's behaviour seems to differ from the standard Nash-equilibrium. An often-overlooked implication of social preferences, however, is that participants do not necessarily face a public-good *game* when researchers present them with a situation whose monetary payoffs have a public-good structure (or when life presents them with a situation that has a public-good structure in terms of money or time costs, for that matter).<sup>1</sup>

Even so, it remains unclear whether the difference between behaviour and the standard Nash-equilibrium is due to a misspecification of the players' preferences (who, *e.g.*, may take others' payoffs into consideration), a mistaken account of the strategic aspects of the interaction, or both. As a consequence, a good account of behaviour in such situations is still missing. However, society's responses to the public-good situations crucially will depend on our understanding of when an agent will choose to contribute. At an abstract level, this paper contributes to such an understanding.

The paper addresses the question of whether a Nash-concept can predict behaviour in one-shot public-good experiments out of sample, once the Nash-concept is based on appropriate measurements of people's preferences. The answer is positive. This is surprising on a number of accounts. First, many researchers tend to understand Nash-equilibrium only as a long-run prediction, not a prediction for one-shot situations. Second, the equilibrium's pre-conditions are missing: in particular, participants do not know their interaction partners' preferences. And third, prior research seemed to suggest that the missing knowledge of others' preferences indeed prevents a successful prediction of behaviour.

Putting the findings of this paper into a broader perspective, I show that the positive-contributions equilibria identified in Wolff (2017) are meaningful for behaviour. In this light, the substantial degree of cooperation in human everyday interactions becomes less surprising. However, the low rates of contributions that we typically observe at the end of repeated public-good experiments do become more surprising. If there are equilibria with substantial contribution levels, and if these equilibria are predictive in one-shot games, why do participants in repeated settings not seem to be able to select a cooperative equilibrium more often?

---

<sup>1</sup>The difference perhaps is seen most easily for highly inequality-averse agents à la Fehr and Schmidt (1999): for them, the typical public-good experiment is a coordination game (with any vector of equal contributions being a pure-strategy equilibrium).

## 1 INTRODUCTION

My analysis of equilibrium selection at the end of this study provides a tentative answer. Once there are multiple equilibria, participants do not agree on the equilibrium-selection criterion to use in my one-shot experiment. It is highly likely that this finding carries over to initial play under repeated settings. In repeated settings, multiple equilibria are even more prevalent. This would explain heterogeneous, non-equilibrium behaviour in initial rounds of repeated games. And from there, the dynamics described in Fischbacher and Gächter (2010) will take over, leading to the observed low long-run contribution levels. On the other hand, the findings suggest that if actors—be it in the lab or outside—only could coordinate on an equilibrium-selection criterion beforehand, many public-good problems potentially could be solved.

Having talked about the broad picture, let me provide a little more detail:

**Equilibrium solutions.** It is well-known that social preferences play a role for behaviour, both in public-good situations and beyond. For example, many ultimatum-game responders decline low offers. Or, for an example that is more specific to this paper, when last-movers have to decide on their contribution in a sequential public-good situation, many of them reciprocate high contribution levels of others.

For all of these situations, it is clear what a Nash-equilibrium looks like for payoff-maximizing agents. Social-preference models have been introduced to provide a Nash-equilibrium solution also for agents who hold particular predefined types of social preferences (*e.g.*, Levine, 1998; Fehr and Schmidt, 1999). But what would the “standard game-theoretic solution” be when taking into account participants’ actual preferences? One of the possible answers is what I call the “revealed-preference Nash-equilibrium” (RPNE). The RPNE is the set of Nash-equilibria that results when measured conditional-contribution preferences are interpreted as best-reply correspondences.

The RPNE’s logic is simple. Measured conditional contributions are how a participant reacts to each possible contribution (vector) of her fellow group member(s), when the other player(s) already has/have made their choice(s) in a sequential public-good situation. If these conditional contributions are taken to be direct expressions of how the participant wants to respond to the respective contribution levels, then conditional contributions are also the best-replies to these contribution levels. In turn, an RPNE is a situation in which the players’ contributions are mutual best-replies (or ‘mutual conditional contributions’). Thus, the RPNE rests on the assumption that what a player prefers to give in response to a contribution vector  $\mathbf{x}$  in a sequential situation is the same as what the player would prefer to give in a simultaneous situation in which she was certain that others will be choosing  $\mathbf{x}$ .<sup>2</sup>

<sup>2</sup>Prior research supports this assumption (Fischbacher et al., 2012).

## 1 INTRODUCTION

**Nash-predictions for one-shot public goods?** In contrast to my paper, most research on social-preference equilibria in public-good situations has started from the understanding that we cannot expect a Nash-concept to predict one-shot behaviour well. Thus, prior research typically has focused on (last-round) behaviour in repeated games. Arguably the most important reason for why a Nash-concept may not be suited to behaviour in one-shot situations is that people would not know others' preference types. Therefore, it would be impossible for them to know the equilibria of the game. Indeed, Healy (2011) finds that “[t]he failure of Nash equilibrium stems in a large part from the failure of subjects to agree on the game they are playing.” While it undoubtedly is true that experimental participants do not know their co-players' true preferences, this paper challenges the notion that a Nash-concept cannot predict one-shot public-good behaviour well.

**Research question 1.** The most informative test of whether a given explanation is meaningful or whether a model simply accommodates the data by virtue of its number of free parameters are quantitative predictions about a specific new situation.<sup>3</sup> Unfortunately, few popular social-preference models come with a calibration that would allow to make such a prediction.<sup>4</sup> The first contribution of this paper is to examine the *predictive* power of RPNE, with the corresponding research question:

**RQ 1.** Can a Nash-concept predict behaviour (even) in one-shot public-good experiments, when it is based on a measurement of preferences in a different sample (and for a substantial part of the data, in a different student population)?

The answer is yes. In particular, I show that the RPNEs calculated in Wolff (2017) are predictive for behaviour in eight different data sets, six of them stemming from earlier studies (Blanco et al., 2011; Guala et al., 2013; Kamei, 2016, for two-player games, and Cubitt et al., 2001; Drouvelis et al., 2015; Dufwenberg et al., 2011, for three-player games).<sup>5</sup>

**Putting the finding into context.** This answer is surprising on two accounts. First, earlier related work by Healy (2011) or Brunner et al. (2021) suggested that

---

<sup>3</sup>Relatedly, Andreoni and Samuelson (2006) call for a focus on predictions pointing out that “[t]he difficulty in interpreting such models is distinguishing when we have uncovered a robust feature of behavior and when we have fortuitously constructed preferences that happen to match some experimental observations.”

<sup>4</sup>Arifovic and Ledyard (2012), Fehr and Schmidt (1999) and Levine (1998) are notable exceptions.

<sup>5</sup>Wolff (2017) categorised the equilibrium sets to be expected in a well-mixed population, contrasting the result to the prediction of the calibrated model by Fehr and Schmidt (1999), with no reference to actual behaviour.

## 1 INTRODUCTION

a Nash-concept based on preference-measurements does not account for public-good behaviour. The reason for the differing finding may be that the RPNE approach implicitly incorporates reciprocity concerns, a feature that is absent in both Healy (2011) and Brunner et al. (2021) but that arguably is important for behaviour in public-good situations. On top, Fischbacher and Gächter's (2010) results suggest that while participants generally best-respond to their beliefs, their beliefs are not equilibrium beliefs. Note, however, that Fischbacher and Gächter's focus is on repeated interactions, which increases the prevalence of multiple-equilibrium situations. Multiple-equilibrium situations in turn bring about miscoordination because, as we shall see, people differ in the equilibrium-selection criteria they use.

**Research Question 2.** The second reason for why the RPNE's predictive power is surprising *a priori* is that, following the discussion above, participants in seven out of the eight predicted samples do not have any information on their co-players' preferences, so that the common-knowledge-of-preferences assumption is violated. This immediately leads to my second research question:

**RQ 2.** Does incomplete information about preferences (not) play a role?

To answer **RQ 2**, I conduct an additional 'PUBLIC PREFERENCES' experiment that creates an environment that approximates mutual knowledge of preferences.<sup>6</sup> The RPNE's out-of-sample mean squared prediction error is even lower for the PUBLIC-PREFERENCES experiment compared to the other seven data sets. This shows that the violation of the common-knowledge-of-preferences assumption in the 'standard' data sets does compromise the RPNE's predictive power to a certain degree. Furthermore, partitioning the sample into participants for whom core Nash-assumptions are fulfilled *versus* those for whom the assumptions are violated shows that behaviour can be predicted the better, the more closely the assumptions are fulfilled. Arguably, these observations lend support to the idea that the RPNE predicts behaviour for the right reasons.

**Strategic uncertainty, the mechanism, and Research Question 3.** In the remainder of the analysis, I accomplish three goals. First, relating to a discussion in the current prisoner's-dilemma literature, I look at strategic uncertainty. I find that the out-of-sample predictive power of the RPNE calculated in Wolff (2017) for a new 'STANDARD' data set gathered for this paper is much better for those whose elicited beliefs show a low degree of strategic uncertainty. Second, an

---

<sup>6</sup>I will be explicit below about how I deal with the potential signalling incentives, at the same time avoiding 'bad surprises' on the part of the participants (that an action they thought would remain anonymous gets revealed to others); see the two paragraphs just before Section 3.1. In essence, I use a similar approach as Brunner et al. (2021).

## 2 CLOSELY RELATED LITERATURE

individual-level analysis of the PUBLIC-PREFERENCES data supports the findings from the main part: participants' behaviour can be predicted surprisingly well, and the better, the more the equilibrium pre-conditions tend to hold. Finally, the individual-level analysis allows to look at a third research question:

**RQ 3.** Which equilibrium will be selected in case of multiple equilibria?

As posited by Fehr and Schmidt (1999), the modal choice corresponds to the Pareto-dominant equilibrium. However, it accounts for only 36% of the choices when participants face multiple equilibria. The majority of the participants seem to disagree about how to solve the equilibrium-selection problem: some choose the 'average' equilibrium (the one with the average sum of contributions), some choose the 'most pessimistic' equilibrium, and 36% choose non-RPNE actions. Among the participants for whom core RPNE assumptions seem to be fulfilled, the percentage of non-RPNE choices goes down to 19%, while the relative frequencies of the different equilibrium choices are comparable.

## 2 Closely related literature

During the long history of public-good research, there have been a large number of studies aiming at understanding public-good contributions through participants' measured preferences and their beliefs (*e.g.*, Offerman et al., 1996, for an early example). In the context of this study, important contributions in this tradition are Fischbacher and Gächter (2010) and Fischbacher et al. (2012), as they also rely on conditional-contribution preferences.

In their study of a finitely-repeated public-good situation, Fischbacher and Gächter (2010) suggest that participants generally best-respond to their beliefs (judging by their elicited conditional-contribution preferences), but that their beliefs are not equilibrium beliefs (and participants update the beliefs suboptimally). Fischbacher et al. (2012) establish the behavioural validity of conditional-contribution preference measurements for actual public-good play even more forcefully. These results would suggest that it is the strategic-interaction aspect that would be the most likely culprit if behaviour deviates from a preference-based Nash-prediction.

In contrast, Ambrus and Pathak (2011) promote the idea that participants of finitely-repeated public-good experiments actually are playing an equilibrium. However, they restrict their focus explicitly to "repeated games in which players are experienced," "[t]o approximate the complete information assumption of our model." The statement clearly implies that the complete-information assumption of their Nash-equilibrium approach (or mine) may be violated in one-shot situa-

## 2 CLOSELY RELATED LITERATURE

tions such as those in the data sets I study. A study by Healy (2011) shows that this indeed is the case.

Healy (2011) and Brunner et al. (2021) both measure distributional preferences to make an elicited-preference-based Nash-prediction in normal-form  $2 \times 2$  games. Healy (2011) examines the conditions that Aumann and Brandenburger (1995) identify as sufficient conditions for a Nash-equilibrium. He concludes that Nash-equilibrium fails to predict behaviour predominantly because participants correctly predict how their opponent would rank the four possible outcomes of a particular game in only 64% of the games.

Brunner et al. (2021) inform their participants about their opponents' elicited preferences in one treatment (similar to my PUBLIC-PREFERENCES experiment). They compare the Nash-equilibrium's predictive power to a treatment without this information and find a significant increase in the amount of equilibrium play: the display of the opponent's preferences increases the percentage of equilibrium play from some 42-47% to some 51-52%—in their  $2 \times 2$  games. Comparing these figures to a random benchmark of 50%, it seems safe to say that the equilibrium does not seem to be a very good predictor of behaviour.

Let me now turn to the models I will be using for prediction. At the focus of this study is the 'revealed-preference Nash-equilibrium' (RPNE) introduced in Wolff (2017). In that paper, the concept is presented, and the sets of equilibria that would arise in a well-mixed population are categorised. The categorisation is done for a three-player situation with a marginal per-capita return  $\mu$  of  $\mu = 0.5$ , and for two-player situations with  $\mu = 2/3$  and  $\mu = 0.75$ . Finally, Wolff (2017) compares how often different equilibrium-set types would occur under the different parameter combinations to the predictions for the calibrated model of Fehr and Schmidt (1999). The general upshot is that the RPNE predicts positive contributions substantially more often than Fehr and Schmidt (1999, *e.g.*, in 38% as opposed to 6% of the cases for the three-player setting). What we do not learn from that paper is how either model performs in predicting actual behaviour.

Next to the model of Fehr and Schmidt (1999) and the 'selfish Nash-prediction', I am aware of two calibrated models in the literature that would be applicable to one-shot public-good situations like the ones I study. In an early social-preference model, Levine (1998) posits that others' utility enters a players' own utility function with a higher weight, the more the player thinks that these others are of an altruistic type. Levine's basic assumption—that players know only the distribution of types in the population—is likely to be much closer to the experimental conditions in most of the data sets I study than the common-knowledge-of-preferences assumption in the other models. However, its predictions coincide with the 'selfish Nash-prediction' in all experiments I study (note that players cannot update their beliefs about the opponents' type in a simultaneous game, and the calibrated model is such that the population's average type is slightly



spiteful). Given what we know from the literature, this prediction does not correspond well with actual data.

In contrast to the ‘selfish Nash-equilibrium’, Fehr and Schmidt (1999), and Levine (1998), Arifovic and Ledyard (2012) present a model that is tailored specifically to public-good situations. In essence, Arifovic and Ledyard combine outcome-based social preferences with heterogeneous types with a kind of “reactive-learning” model (as opposed to strategic behaviour). However, the learning part does not apply to my one-shot setting, which is why I only consider the social-preference part of their model which is meant to account for unexperienced play. Arifovic and Ledyard show that the general versions of their model and the earlier models of Fehr and Schmidt (1999) and Charness and Rabin (2002) are equivalent, but that the models differ in terms of the imposed parameter restrictions. The parameter restrictions then produce differing predictions. Most importantly, the model of Arifovic and Ledyard (2012) is able to account for contributions that are neither 0 nor participants’ full endowment.

Finally, research question **RQ 2** parallels current discussions in the literature on indefinitely-repeated prisoners’-dilemma experiments. In particular, Vespa et al. (2021) analyse (and document) in depth the role of strategic uncertainty, while Kartal and Müller (2021) focus on the importance of the *incomplete* information about the opponent’s preferences. The findings of the current paper nicely complement these findings by showing that both, strategic uncertainty and the degree of knowledge of others’ preferences play an important role also in one-shot public-good situations. Note that the difference between a prisoners’ dilemma and a linear public good is non-trivial, as behaviour in prisoners’ dilemmas by construction cannot be as rich as that in public-good games. In particular, in a prisoners’ dilemma, there cannot be any imperfect conditional cooperators or triangle contributors, two types that have been identified robustly in the public-good literature—and one of which has been identified by Fischbacher and Gächter (2010) as an important ingredient of the explanation of contribution decay in repeated public-good experiments.

### 3 The Data

In this paper, I use the data from eight data sets. Six of the data sets are from earlier studies that contained one-shot simultaneous linear public-good situations with two (Blanco et al., 2011; Guala et al., 2013; Kamei, 2016) or three players (Cubitt et al., 2001; Drouvelis et al., 2015; Dufwenberg et al., 2011).<sup>7</sup> The three-

---

<sup>7</sup>It was unexpectedly hard to find plain-vanilla two- or three-player simultaneous-public-good experiments that were played without repetition and without any institutions (such as punishment, reward, pre-play communication, etc.) but with multiple contribution levels (*i.e.*,

### 3 THE DATA

player studies all used marginal per capita returns  $\mu = 0.5$ , while the two-player studies had different  $\mu$ s (0.7, 0.75, and 0.6, respectively). To these data sets, I add two additional experiments that I call STANDARD and PUBLIC PREFERENCES.

The STANDARD Experiment is a standard one-shot simultaneous linear public-good experiment, which had  $\mu = 2/3$ , contribution levels of  $\{0, 3, 6, \dots, 15\}$  “guilders” (2 “guilders” = 1 Euro). After choosing their contribution to the public good, participants had to report their belief on what percentages of other players had chosen each possible contribution level. Their payment would be 20 guilders in case the sum of percentage-point deviations of their belief from the actual percentages would not be larger than five percentage points.<sup>8</sup>

The PUBLIC-PREFERENCES Experiment is more complicated. It consists of seven parts, one of which is drawn randomly for payment. For none of the experimental parts do participants get any direct feedback before the end of the session.

**svo** A social-value orientation task similar to the one presented in Murphy et al. (2011). Used to calculate individual-level Fehr-Schmidt- and Arifovic-Ledyard-predictions in Section 4.2.

**PREFS1** A standard elicitation of conditional-contribution preferences (“PREFS task”, Fischbacher et al., 2001), detailed in Section 3.1.

**PREFS2 + BELIEFS.** Repetition of the PREFS task with a new opponent. Then, I elicited beliefs on the expected first-mover contribution, to train participants in the elicitation method used in SIMPGBELIEFS: probabilistic beliefs elicited by a binarised scoring rule (McKelvey and Page, 1990; Hossain and Okui, 2013, probability of receiving a prize of 2 Euros determined by a quadratic scoring rule; I do not analyse the beliefs from this part).<sup>9</sup>

**SIMPG.** The focal simultaneous public-good interaction also detailed in Section 3.1.

---

that would go beyond a prisoners’-dilemma setting). I first asked for pointers via the ‘ESA-discuss’ e-mail list and got a substantial number of replies; unfortunately, most of them turned out to be unsuited for the purposes of this paper. I then checked the Cooperation Databank (Spadaro et al., 2020) and found a number of papers, out of which, however, some of the matches were unsuitable, too (*e.g.*, because they examined sequential-play setups or non-student samples), or I simply was not able to obtain the data.

<sup>8</sup>More precisely, the sessions would consist of two parts, one of which would be drawn randomly to be payoff-relevant. Part 1 was the public-good situation, whereas Part 2 consisted of the belief-elicitation above plus a completely unrelated experimental task. Each task was described to participants only after completing the preceding task.

<sup>9</sup>Note that by the transformation of payoffs into lottery tickets, the binarised scoring rule is proper under any expected-utility risk preferences, and even for non-expected-utility agents whose preferences satisfy a mild monotonicity condition (*cf.* Hossain and Okui, 2013).

### 3 THE DATA

**SIMPGBELIEFS.** Elicitation of beliefs on the likelihood of the interaction partner choosing each possible action in the SIMPG part (binarised scoring rule with payoffs of 20 Euros if successful and 4 Euros if not successful).

**STABILITYBELIEFS.** Elicitation of beliefs in others' elicited-preference stability, with respect to the SIMPG-opponent and three randomly-chosen others. Participants saw the respective other participant's response vector from the PREFERENCES1 part. Then, they had to state a probabilistic belief on the response-vector of the same other participant from the PREFERENCES2 part. To be exact, participants had to state for each possible first-mover contribution how likely it was that the other person chose each of their possible contribution levels in PREFERENCES2 (that is, they had to specify  $6 \times 6$  probabilities for each of the four others). For each of the four others whose behavioural stability participants had to assess, one first-mover contribution was randomly drawn. Participants were paid by a binarised scoring rule for their belief accuracy in the four randomly-drawn cases, with a prize of 6 Euros per lottery.

**PREFERENCES3.** Final repetition of the PREFERENCES task with a new interaction partner.<sup>10</sup>

The focus of the PUBLIC-PREFERENCES Experiment is on the predictability of contribution behaviour in an environment that aims to approximate the pre-conditions for an RPNE—the SIMPG part—and how this predictability depends on whether these pre-conditions are fulfilled. The SIMPG part is a standard two-player one-shot linear public-good experiment, except for the fact that participants see their interaction partner's responses from the PREFERENCES1 part. I assess the individual-level predictability of participants' contribution behaviour in Section 4.2 by contrasting the SIMPG-part choices to the RPNE predictions that result from the PREFERENCES1 measurements.

I study two pre-conditions for an RPNE: (i) that participants' elicited conditional-contribution preferences are stable in the sense that they do not change every time I elicit them; and (ii) that the induction of mutual knowledge of conditional-contribution preferences is successful. To assess pre-condition (i), I elicit participants' preferences for conditional cooperation three times within a session: twice at the beginning, and a third time as the final part of the session (PREFERENCES1, PREFERENCES2, and PREFERENCES3).<sup>11</sup> And to assess pre-condition (ii), the STABILITYBELIEFS

---

<sup>10</sup>In contrast to the first two PREFERENCES tasks, the first-mover in PREFERENCES3 was shown the response-vector of the second-mover from the PREFERENCES1 part before deciding on her (unconditional) contribution. However, the situation of the second-mover was exactly the same as in the PREFERENCES1 and PREFERENCES2 parts. For the purpose of this paper, I therefore regard the PREFERENCES3 part simply as a second repeat-measurement of participants' preferences. I did not analyse the PREFERENCES3 first-mover behaviour.

<sup>11</sup>To make the repeated elicitation of preferences more natural, participants are always matched to a new other player after each part.

### 3 THE DATA

Experiment elicits participants' probabilistic beliefs about others' responses in the Prefs2-Experiment showing them these others' responses from the Prefs1-Experiment.

**Note on player-type categories in PUBLIC PREFERENCES.** Relating to the pre-conditions mentioned above, I categorise participants into  $2 \times 2$  categories. In the following paragraphs, I outline the categories and specify the corresponding criteria. I categorise all participants as having "consolidated preferences" whose average squared difference from the mean response to each first-mover contribution across Prefs1, Prefs2, and Prefs3 is at most 2. This criterion would be fulfilled with equality if a participant replies to each first-mover contribution the same way twice, deviating on the third occasion by one increment of 3 Euros in all contingencies.<sup>12</sup> Participants who violate the criterion are categorised as having "floating preferences". I choose these labels to represent the (lack of) volatility in responses without referring to any specific model.

In relation to pre-condition (ii), a participant is categorised as having *incomplete information* with respect to others' preferences or conforming to *mutual knowledge* (of preferences) based on her STABILITYBELIEFS. In the STABILITYBELIEFS part, each participant sees the choices of four other participants from the Prefs1 part and has to state a probabilistic belief about the four others' choices in the Prefs2 part. For the *incomplete-information/mutual-knowledge* categorisation, I focus on the participant's beliefs about the three players who were *not* the participant's SIMPG-opponent. I do so to show that the categories are characteristics of the *person* rather than specific to the situation.<sup>13</sup> I categorise a participant as a *mutual-knowledge* type if she places at least 80% probability on the three other players responding to all possible first-mover contributions the same way in the Prefs1- and the Prefs2-experiments, and as an *incomplete-information* type, otherwise.<sup>14</sup>

The above typology partitions the population into four groups with the fol-

---

<sup>12</sup>Using this criterion, there are 66 *approximately stable* participants (out of 152). If we were to use a median split instead, the threshold would almost double, to 11/3. Only eight additional participants have an average squared difference from the mean response of less than 11/3, so that the results would not differ very much.

<sup>13</sup>The predictive power actually is slightly worse when categorising participants by the SIMPG-opponent's expected stability (with a category-wise-weighted mean squared prediction error of 0.0087 instead of 0.0070). This is consistent with a person-specific characteristic that predicts the expected stability of the SIMPG-opponent as well as the participant's behavioural consistency with the RPNE. The additional noise from relying on a single stability-belief measurement seems to be (slightly) larger than the decrease in noise associated with the actual-interaction-specific measurement. Having said this, the interaction-specific characteristics will be important in the section on individual-level predictions.

<sup>14</sup>Changing the threshold to, e.g., 70% does not change the results in any meaningful way.

### 3 THE DATA

lowing relative frequencies: *consolidated preferences/mutual knowledge*: 30%; *floating preferences/mutual knowledge*: 13%; *consolidated preferences/incomplete information*: 26%; and *floating preferences/incomplete information*: 31%.

#### **Note on signalling incentives in the PUBLIC-PREFERENCES Experiment.**

Note that if participants know that their behaviour in one experiment may be revealed to others in the next experiment, they may have potential signalling incentives in the first of the two.<sup>15</sup> If such signalling incentives changed participants' Prefs1-responses—or lead to beliefs that others change their Prefs1-responses—elementary pre-conditions for an RPNE would be violated. If only the responses were affected, the out-of-sample predictions would be unaffected, but individual-level predictions would suffer. If the beliefs were affected(, too), that would imply a failure of the induction of mutual knowledge of preferences. In any case, if at all, the RPNE's predictive power would be worsened. In that sense, if the signalling incentives were effective, my paper would provide a conservative estimate of the explanatory power of the RPNE.

Having said this, my experimental design addresses the signalling problem through a number of design choices (discussed in full detail in Wolff, 2015, on a very similar earlier design; see also Brunner et al., 2021, for a similar approach). Most importantly, participants make decisions in seven distinct experimental parts with new interaction partners in each of them, being paid for only one randomly chosen experiment (which should make signalling prohibitively costly). They do not get any information about others' behaviour before the SIMPG-experiment, and each experiment is explained only as soon as it begins. While it is impossible-in-principle to show there have been no signalling attempts by participants, I could not find any evidence of signalling in the data.

#### **3.1 The SIMPG- and the Prefs-experiments**

The SIMPG-experiment consists of a simultaneous two-player linear public-good situation with an  $MPCR = \frac{2}{3}$  and an endowment of 15 Euros. Each player has to choose a contribution to the public good from the set  $\{0, 3, 6, 9, 12, 15\}$  Euros, which is multiplied by  $\frac{4}{3}$  and divided equally among the two players, regardless of each player's own contribution. In addition, players see the elicited Prefs1-preferences of their opponent before making their choice.

In the Prefs-experiment, participants face the same two-player linear public-good payoff structure with an  $MPCR = \frac{2}{3}$  and an endowment of 15 Euros as in the SIMPG-experiment. However, the Prefs-experiment differs from the SIMPG in

---

<sup>15</sup>To avoid deceiving participants, the instructions included the sentence that “your behaviour from one of the earlier parts will possibly be displayed to other participants in a later part.”

### 3 THE DATA

that there is no information on the other player, and in that the PREFERENCES-experiments are sequential games: one participant moves first and the other moves second, being informed of the first participant's choice. Participants have to decide in either role. First, they specify their first-mover contribution to the public good that is implemented if they are not (randomly) chosen to be the second-moving player. Then, I elicit their second-mover choices using the strategy method: they are presented with all possible first-mover contributions and asked to specify their 'conditional' contributions.<sup>16</sup>

To limit the scope for confusion as a major source of revealed-preference instability, I took three measures. First, I restricted the simultaneous game to a two-player six-action game rather than the usual three- or four-player games with 11-21 actions. While the MPCR may look a little complicated, all game payoffs were integer amounts. Second, I always displayed the full payoff matrix in the relevant parts. Moreover, I highlighted the relevant part of the matrix in the preference-elicitation parts of the PREFERENCES-experiments, so that participants would know exactly what payoff profile each of their actions meant. As a third measure, I recruited *experienced* participants.<sup>17</sup> Participants in the experiment had participated in at least one public-good experiment and at least four additional other experiments, with no upper limits.

## 3.2 Procedures

**The STANDARD Experiment.** The STANDARD Experiment was conducted in April 2021, and thus had to be conducted online. Participants were invited to a virtual meeting room where they could not see each other or communicate with other participants. There, we welcomed participants, checked their identities, and were available for questions via the chat function throughout the experiment. Once we documented that all participants in the virtual room had registered for the experimental session before, we sent out personalized links for the experiment. Participants would open the links, consent to our laboratory rules, and read the experimental instructions. Once all participants had answered all control questions correctly, the experiment would start. Participants earned about 13.80 Euros (USD 16.60) on average for about one hour, including a show-up fee of 5 Euros.

---

<sup>16</sup>The order of the combinations was randomised individually for each player. Responses were elicited one-by-one for two reasons: (i) to make each decision as salient as possible, (ii) to elicit 'smooth' response-patterns only in case preferences gave rise to them.

<sup>17</sup>I nonetheless asked the usual comprehension questions; participants could only proceed to the experiment after answering all questions correctly.

## 4 RESULTS

**The PUBLIC-PREFERENCES Experiment.** On the day of the experiment, participants were welcomed and asked to draw lots in order to assign them to a cabin. There, they would find some explanation on the general structure of the experiment and on the selection of the payoff-relevant experiment (and role, if applicable). The instructions for each experiment were displayed directly on their screen during the corresponding part. The (translated) general and on-screen instructions are gathered in Appendix B.

Participants earned on average 19.33 Euros (USD 22) for about 90 minutes; this includes a 2-Euro flat payment for the completion of a post-experimental questionnaire. Altogether, seven sessions with a total of 152 participants were conducted at the LakeLab of the University of Konstanz. The Prefs1 data of the first four of these seven sessions entered the calculations in Wolff (2017). To have a clean separation, I use only the last three sessions (PUBLIC PREFERENCES-NEW,  $N = 70$ ) for assessing the out-of-sample predictions in Section 4.1. For the individual-level analyses in Section 4.2, I then use the data from all seven sessions (PUBLIC PREFERENCES-ALL).

## 4 Results

I structure the results section into three parts. In the section 4.1, I focus on the out-of-sample predictions. This means calculating population-level predictions for how many participants choose which contribution level, based on calibrations from earlier studies. This part is a test of the different models' external validity and penalizes model( calibration)s for over-fitting.

Section 4.2 examines the mechanism by looking at individual-level predictions. Here, I fit model parameters based on the social-value-orientation task (for the Fehr-Schmidt and Arifovic-Ledyard models) or measure conditional-contribution preferences (for the RPNE) to generate individual predictions for each participant for the one-shot simultaneous public-good situation.

Finally, in Section 4.3, I study participants' equilibrium selection in case of multiple equilibria. For this purpose, I relax the assumption that participants always choose the Pareto-dominant equilibrium, and compare participants' choices in the simultaneous public-good situation to individual RPNE predictions that rely on different equilibrium-selection criteria.

### 4.1 Out-of-sample predictions

Table 1 reports the mean squared prediction errors of the out-of-sample RPNE predictions calculated in Wolff (2017) for the eight data sets. Note that Kamei (2016) and Blanco et al. (2011) use marginal per capita returns ( $\mu = 0.6$  and  $\mu =$

## 4 RESULTS

0.7, respectively) for which I do not have an RPNE prediction. I use the predictions for  $\mu = 2/3$  for these two data sets, arguing that the  $\mu$ s are sufficiently close to yield similar results.<sup>18</sup>

As benchmarks, I also report the prediction errors for the standard Nash-equilibrium with selfish preferences; the calibrated Fehr-Schmidt (1999) model; and the calibrated model by Arifovic and Ledyard (2012). The prediction of the calibrated Levine (1998) model coincides with ‘selfish Nash’. For the RPNE prediction in case of multiple equilibria, I adopted the Pareto-dominance criterion from Fehr and Schmidt (1999).

The table provides two insights. First, the RPNE predicts the data from all of the 2-player data sets best and ties with the calibrated model by Arifovic and Ledyard (2012) for the 3-player data sets (the RPNE predicts one data set better and has a slightly lower weighted mean squared prediction error: 0.0115 vs. 0.0120).<sup>19</sup> The RPNE’s predictive success for the games from the literature is remarkable because it happens despite of a number of slight differences in the setups. First, the MPCRS of two studies are different ( $\mu = 0.7$  for Blanco et al., 2011,  $\mu = 0.6$  for Kamei, 2016) from the data the prediction was based on ( $\mu = 2/3$ ). Second, I had to bin the data from the earlier studies into 6 contribution levels (in the original data, participants could contribute any integer amount between 0 and 10 in Blanco et al. and Guala et al., and between 0 and 20 in the other studies).<sup>20</sup> And third, most of the earlier studies had different treatments. In order not to run the risk of cherry-picking the best-fitting treatments, I simply use the data of all treatments.

The second insight that Table 1 provides is that the predictive power is particularly strong where we would expect it to be strong. First of all, the RPNE’s predictive power is particularly strong for the PUBLIC PREFERENCES-NEW data, where participants ‘know who they are playing against’.<sup>21</sup> Second, the RPNE’s

---

<sup>18</sup>In fact, the comparative statics are exactly what we would expect given the MPCRS: average contributions in Kamei (2016;  $\mu = 0.6$ ) are lowest (30%), followed by the RPNE prediction ( $\mu = 2/3$ , avge: 37%) and those in Blanco *et al.* (2011;  $\mu = 0.7$ , avge: 48%).

<sup>19</sup>Note that the RPNE does not predict worse in the 3-player games compared to the 2-player games. It is the model by Arifovic and Ledyard (2012) that predicts better in the 3-player as compared to the 2-player games (the same holds true for the other two models).

<sup>20</sup>Note also that I pooled all data from the first part of Kamei’s study in which each participant simultaneously interacts in two public-good situations with different opponents. Using only the ‘left’ game or only the ‘right’ game does not change the results in any meaningful way.

<sup>21</sup>Not surprisingly, the results do not differ much if I instead predict the PUBLIC PREFERENCES-ALL data. Note also that, while the effect clearly is there, the mean squared prediction error in the 7th data row of Table 1 slightly exaggerates its strength. As we can see from looking at the mean squared prediction errors of the four subgroups in the last 4 lines of the Table, the small size of the prediction error stems in part from deviations by the individual subgroups setting each other off. To address this issue, we need the individual-level analysis in the following Section.



## 4 RESULTS

predictive power is particularly strong for those participants of the STANDARD treatment who report low strategic uncertainty. To measure subjective strategic uncertainty in the STANDARD treatment, I calculate the sum of squared deviations of the participants' action-beliefs from a uniform distribution. Then, I use a median split to divide the observations into a "high strategic uncertainty" and a "low strategic uncertainty" category.

As we can see from the eighth row of Table 1, the predictive power is relatively low for those whose action-belief is comparatively close to uniformity. In contrast, the predictive power approaches that for the PUBLIC PREFERENCES-NEW treatment for those whose action-belief tends to be focused on a single action of their opponent, as evidenced by the sixth data row of Table 1.<sup>22</sup> The effect is even stronger if we restrict our attention to the quartile of the STANDARD participants who report the least strategic uncertainty (mean squared prediction error: 0.0025). Finally, in the PUBLIC PREFERENCES-NEW treatment, the prediction error is smallest for those for whom the induction of mutual knowledge of preferences seems to work. What is surprising is that the distinction between "consolidated" and "floating" preferences does not seem to matter for the RPNE's predictive power. I will explore the role of the "consolidation" of preferences further in the within-sample individual-level analysis below.

Figure 1 shows a histogramme for the RPNE prediction and the data from the two treatments of this study, to obtain an idea of where the predictions fail. Figure 1 suggests that in STANDARD—where people do not know who they are playing—many who should be contributing nothing 'overplay' by choosing low-to-medium contributions (albeit it is too early to draw definite conclusions because I still refer to aggregate-level data here). This effect is strongly reduced in the PUBLIC PREFERENCES-NEW treatment. In this treatment, there seems to be a (smaller) shift from full-contributions to medium contributions. This suggests that—in contrast to Fehr and Schmidt's (1999) assumption which I also have been following—the relevant equilibrium-selection criterion may not be Pareto-dominance for all of the participants.

So far, I have demonstrated the predictive power of the RPNE concept for two-player public-good situations in out-of-sample (and, mostly, out-of-participant-pool) predictions. I have shown that the concept predicts particularly well for participants whose subjective strategic uncertainty is low, and for participants who generally find the induction of mutual knowledge of preferences in PUBLIC PREFERENCES-NEW credible. Out-of-sample predictions have the great advantage of demonstrating external validity and penalizing over-fitting. On top, they can be tested even when the the assumptions of the model are violated

---

<sup>22</sup>The contrast replicates, albeit not as pronouncedly, in treatment PUBLIC PREFERENCES-NEW, with mean squared prediction errors of 0.0197 vs. 0.0092.

## 4 RESULTS

Data	‘selfish Nash’	FS (1999)	AL (2012)	Wolff (2017)*
Kamei (2016; $n = 2, \mu = 0.6; N = 300$ )	0.0761	0.0484	0.0575	<b>0.0080</b> <sup>†</sup>
Blanco <i>et al.</i> (2011; $n = 2, \mu = 0.7; N = 72$ )	0.1042	0.0653	0.0181	<b>0.0104</b> <sup>†</sup>
Guala <i>et al.</i> (2013; $n = 2, \mu = 0.75; N = 410$ )	0.1384	0.0615	0.0451	<b>0.0200</b>
Cubitt <i>et al.</i> (2011; $n = 3, \mu = 0.5; N = 87$ )	0.0613	0.0374	<b>0.0101</b>	0.0145
Drouvelis <i>et al.</i> (2015; $n = 3, \mu = 0.5; N = 150$ )	0.0736	0.0470	<b>0.0190</b>	0.0192
Dufwenberg <i>et al.</i> (2011; $n = 3, \mu = 0.5; N = 303$ )	0.0454	0.0247	0.0090	<b>0.0068</b>
STANDARD ( $n = 2, \mu = 2/3; N = 72$ )	0.1191	0.0757	0.0269	<b>0.0111</b>
High strategic uncertainty	0.1632	0.1100	0.0468	<b>0.0255</b>
Low strategic uncertainty	0.0824	0.0488	0.0144	<b>0.0043</b>
PUBLIC PREFERENCES-NEW ( $n = 2, \mu = 2/3; N = 70$ )	0.0702	0.0384	0.0139	<b>0.0020</b>
floating preferences, incomplete information	0.1136	0.0716	0.0207	<b>0.0108</b>
consolidated preferences, incomplete information	0.0473	0.0313	0.0240	<b>0.0132</b>
floating preferences, mutual knowledge	0.0567	0.0297	0.0168	<b>0.0035</b>
consolidated preferences, mutual knowledge	0.0761	0.0376	0.0152	<b>0.0022</b>
<i>p</i> -value, Wilcoxon signed-ranks test against Wolff (2017); $N = 8$	0.008	0.008	0.055	–

Table 1: Mean squared prediction errors of the stated models for the different data sets (the prediction of Levine’s, 1998, coincides with ‘selfish Nash’). \*In case of multiplicity, I adopt Fehr & Schmidt’s (1999) Pareto-dominance criterion. <sup>†</sup>Prediction for  $n = 2; \mu = 2/3$ .

as in the STANDARD treatment (much like in the seminal market experiments of Vernon Smith, where participants did not know anything about others’ valuations).

However, we need individual-level (within-sample) analyses to explore whether behaviour reflects the modelled mechanism at least to some degree, and at least when the pre-conditions are approximated. Even more importantly, I need the individual-level analysis to enable me to answer research question **RQ 3**, how participants select their contributions in case of multiple equilibria. This is what the next section explores. Both questions are meaningful only in a PUBLIC-PREFERENCES context, which is why I did not collect conditional-contribution schedules in STANDARD.<sup>23</sup>

### 4.2 Studying the Mechanism: Individual-level predictions

To obtain more information on the mechanism, I conduct an individual-level analysis. In the analysis, I still focus on predicting behaviour in a simultaneous public-good task using elicited conditional-contribution preferences, this time for individual RPNE predictions. The individual-level analysis differs from

<sup>23</sup>Note that the mechanism in STANDARD *needs* to be different from the model mechanism because participants do not know at all whom they are playing with. Thus, it does not make sense to study the mechanism in STANDARD the same way as in the PUBLIC-PREFERENCES context.

## 4 RESULTS

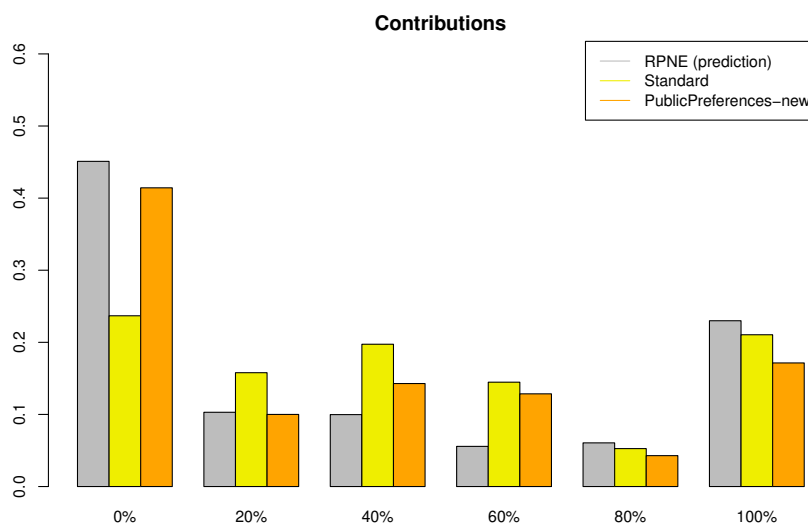


Figure 1: Histogramme for the RPNE prediction from Wolff (2017) and the data from the STANDARD and the PUBLIC PREFERENCES-NEW treatments.

the out-of-sample approach particularly in that, in the individual-level analysis, there is a specific predicted contribution level for each participant.

Table 2 presents the hit rates for the different models. I switch to hit rates because mean squared prediction errors do not allow to address the question for the relevant equilibrium-selection criterion below. If the Pareto-dominant equilibrium prescribes a contribution of 12 and a participant chooses a contribution of 9, then Pareto-dominance does not seem to be the relevant criterion, even if the deviation is only one increment.<sup>24</sup>

To calculate the hit rates in Table 2 for the models by Fehr and Schmidt (1999) and Arifovic and Ledyard (2012), I first estimated participants' individual model parameters from the SVO data using a maximum-likelihood algorithm. For the RPNE prediction, I used participants' own Prefs1-choices together with their actual opponent's Prefs1-choices. I again adopted Fehr & Schmidt's (1999) Pareto-dominance criterion and ignored all cases in which the model is mute (because the RPNE set is empty).<sup>25</sup>

<sup>24</sup>The main qualitative results continue to hold when I use mean squared prediction errors instead.

<sup>25</sup>Counting these cases as 'misses' would yield the following percentages: 48, 34, 45, 46, and 65, in the order given in Table 2). This is a lower bound for the true hit rate because the model is incomplete: the model cannot be assumed to predict that participants do not make any choice at all (which is the implicit assumption in what I referred to as lower-bound hit rates). Possible alternatives may be to prescribe random or 'selfish-Nash' behaviour in case of an empty RPNE set.

## 4 RESULTS

	‘selfish Nash’	FS (1999)	AL (2012)	RPNE*
PUBLIC PREFERENCES-ALL ( $N = 152$ )	41	41	30	51
floating preferences, incomplete information	20	23	19	36
consolidated preferences, incomplete information	40	40	40	45
floating preferences, mutual knowledge	47	49	23	53
consolidated preferences, mutual knowledge	57	54	41	65
$p$ -value of regression coefficients, baseline: RPNE ( $N = 4$ )	0.071	0.085	0.002	-

Table 2: Hit rates for individual-level predictions of the stated models for the PUBLIC PREFERENCES-ALL data (in %). \*In case of multiplicity, I adopt Fehr & Schmidt’s (1999) Pareto-dominance criterion. Cases in which the RPNE set is empty are excluded. The (dummy-)regression regresses the hit-rate percentages of the four types of participants on the model and the type.

Table 2 shows a number of things. First, all of the models clearly are better than a uniform-randomisation heuristic in predicting choices in all the subsets. Having said that, neither the individual Fehr-Schmidt prediction nor the individual Arifovic-Ledyard prediction offer any improvement over the ‘selfish-Nash’ prediction. Recall, however, that both models were better at predicting the aggregate data on all eight data-sets in Table 1, with the Arifovic-Ledyard prediction always being ‘ahead’ of the Fehr-Schmidt prediction. This discrepancy between aggregate-level and individual-level fit echoes the findings of Blanco et al. (2011) and shows that they also apply (and more forcefully so) to the model of Arifovic and Ledyard (2012).

Finally, Table 2 shows that the RPNE model does better in predicting individual behaviour than the other models for all subsets of the data. We further observe that also on the individual level, the RPNE predicts better the better its preconditions seem to be fulfilled. For the subset of participants whose preferences seem to be ‘consolidated’ and who generally think that the PREFERENCES1-responses reflect others’ preferences, the (Pareto-dominant) RPNE exactly predicts about two thirds of all choices.

Intriguingly, when looking at the subsets of participants, the important dimension again seems to be that of whether participants believe they are in a ‘mutual-knowledge-of-preferences world’. As in the out-of-sample predictions we see also in the individual-level predictions that the improvements in predictive power are always much larger going from an ‘incomplete-information’ category to the matched ‘mutual-knowledge’ category than going from a ‘floating-’ to the matched ‘consolidated-preferences’ category.

Before we turn to an analysis of equilibrium selection, let me briefly look at the mechanism behind the findings. Is it that different subsets of people believe in equilibrium to different degrees or do they respond to their own beliefs to

## 4 RESULTS

different degrees? The answer seems to be a combination of both.

In terms of the aggregate probabilities that participants put on the event that their opponent plays according to (one of) the RPNE action(s), there is a difference in the averages. Participants with ‘incomplete information’ place on average 34% probability on RPNE play by their opponent if they have ‘floating preferences’ and 53% if they have ‘consolidated preferences’. For participants who act under ‘mutual knowledge’, the according figures are 50% for the ‘floating-preference’ type and 72% for those with ‘consolidated preferences’.<sup>26</sup>

The obvious next question would be to what degree participants act on the given beliefs. Unfortunately, a direct analysis of best-response rates is unreliable because we do not know participants’ best-responses to non-degenerate beliefs, and most beliefs are mixed. To obtain at least a somewhat robust rough measure, I consider an action to be an ‘approximated best-response’ if it is the PREFERENCES1-response to any of: the belief mode, the average belief rounded to the next-possible value or the average belief rounded down to the next-possible value (to allow for some pessimism).

Using this measure, contributions are ‘approximated best-responses’ in 43% (floating preferences, incomplete information), 50% (consolidated preferences, incomplete information), 64% (floating preferences, mutual knowledge), and 78% (consolidated preferences, mutual knowledge) of the cases.<sup>27</sup> Judging by this—admittedly crude—measurement, the question of whether participants feel they are in a ‘mutual-knowledge’-approximating environment again seems to be more important than whether participants have ‘consolidated preferences’.

### 4.3 Equilibrium selection

**RQ 3** poses the question of what equilibrium—if any—participants will select in case of multiple RPNE. About one third of the PUBLIC-PREFERENCES participants face an RPNE set that has at least two elements. For the predictions in the preceding Sections, I adopted Fehr and Schmidt’s (1999) Pareto-criterion, selecting the RPNE that would yield the highest payoff sum to the pair. But does this assumption correspond to what participants choose? Table 3 gives an answer.

As we can see from the first row of Table 3, the Pareto-criterion is clearly the modal criterion for choices that are consistent with an RPNE prediction. Still,

---

<sup>26</sup>Pair-wise Wilcoxon-Mann-Whitney tests all yield  $p \leq 0.04$  except for the comparison between the ‘intermediate categories’ ( $p = 0.718$ ).

<sup>27</sup>Boschloo-tests yield  $p < 0.05$  for the comparisons between both incomplete-information types and the consolidated-preferences/mutual-knowledge type, as well as between the two floating-preferences types, and  $p \geq 0.173$  for all other comparisons. Note that the finding is only very partially a consequence of certain types having degenerate beliefs and others not: if we exclude the 21 people with degenerate beliefs, the figures change to: 42%, 50%, 55%, and 74%.

## 4 RESULTS

	Pareto	minimum or average	Other	Non-equilibrium
PUBLIC PREFERENCES-ALL ( $N = 50$ )	36	26	2	36
Both incomplete-information types ( $N = 18$ )	28	12	11	50
floating preferences, mutual knowledge ( $N = 16$ )	38	25	0	38
consolidated preferences, mutual knowledge ( $N = 16$ )	44	38	0	19

Table 3: Percentages of choices that correspond to the individual-level RPNE-predictions selected by the criteria given in the column titles, out of all choices under RPNE-sets with at least two elements (in %). For consolidated-preferences/incomplete-information, there were only 3 observations (1 “Pareto”, 2 non-equilibrium). I thus combine the incomplete-information categories.

they make up for only about one third of all choices under multiplicity of equilibria. Another quarter of all choices under multiplicity of equilibria is split among the most pessimistic minimum- and the average-contribution-sum equilibria, roughly in equal parts (the “equal-parts” statement holds for all four categories of participants). Other criteria are hardly ever used, but more than a third of all choices are non-equilibrium choices.

Splitting the above figures up into the participant types I have been using throughout this Section, I obtain a similar picture to what I observed for the predictions: half of the choices by participants who clearly violate the ‘mutual-knowledge’ assumption are non-equilibrium choices, which is true for only one fifth of the choices by participants for whom both equilibrium pre-conditions (I study) seem to be fulfilled.<sup>28</sup> This suggests two things: first, that unsuccessful predictions are not because a majority of participants are using a different equilibrium-selection criterion. While this is true for a non-negligible part of the participants, this part by no means is a majority.

Second, we once more get the impression that the ‘mutual-knowledge’ assumption is the more critical pre-condition: the percentage of non-equilibrium choices increases from 19% for the consolidated-preferences/mutual-knowledge category to 38% if I ‘take away’ the ‘mutual-knowledge’ assumption, but to (unreliable) 67% if I instead ‘take away’ the ‘consolidated-preferences’ assumption. While this observation has to be taken with even more caution than the similar observations above—in particular because I am dealing with different subpopulations here—it fits into the broader picture. I will discuss this picture and suggest an explanation in the following concluding Section.

---

<sup>28</sup>Boschloo-tests comparing the number of non-equilibrium choices between types yield  $p = 0.072$  for the incomplete-information types *versus* the consolidated-preferences/mutual-knowledge type, and  $p \geq 0.353$  for the other two comparisons.

## 5 Summary and conclusion

In this paper, I study whether a Nash-equilibrium based on elicited conditional-contribution preferences (‘revealed-preference Nash-equilibrium’, or RPNE) is able to predict behaviour in one-shot public-good experiments. Both prior research (Healy, 2011; Brunner et al., 2021) and plausibility considerations (participants cannot know each others’ preferences in a one-shot environment) would have cast serious doubt on this endeavour *a priori*. Nonetheless, I show that the RPNE predicts behaviour from six data sets from the literature surprisingly well.

I next report on two additional experiments to test how the RPNE’s predictive power reacts to changes in strategic uncertainty (in the STANDARD experiment), and to changes in the degree to which two of its assumptions are given (in the PUBLIC-PREFERENCES experiment). The PUBLIC-PREFERENCES experiment tests the following assumptions: (i) elicited conditional-contribution preferences are reliable (measured in terms of their test-retest consistency), and (ii) preferences are ‘mutually known’ after a display of the opponent’s elicited conditional-contribution preferences. Accordingly, I divide participants into participants with ‘consolidated’ (*i.e.*, test-retest-consistent) or ‘floating’ (test-retest-inconsistent) preferences, and into participants who are in a ‘mutual-knowledge’ environment or an ‘incomplete-information’ environment with respect to others’ preferences.

The tests yield the following results. First, the RPNE predicts behaviour better the less strategic uncertainty participants express in their elicited beliefs. Second, the RPNE predicts best (in PUBLIC PREFERENCES) if both considered pre-conditions are given: if participants show ‘consolidated preferences’ and believe they are acting in a ‘mutual-knowledge’ environment. This suggests that the RPNE predicts behaviour for the right reasons. Third, the ‘mutual-knowledge’ assumption seems to be more critical in our data-set, prompting the conclusion that the elicited preferences may be more reliable than what the test-retest stability suggests. Note that all of these conclusions are based on out-of-sample RPNE predictions. This suggests that the findings are more robust, but also that they are less informative about the mechanism.

To obtain more information on the mechanism, I conduct an individual-level analysis. The analysis still focuses on predicting behaviour in a simultaneous public-good task, but this time I use the participants’ own elicited conditional-contribution preferences for individual RPNE predictions. The individual-level analysis differs from the out-of-sample approach particularly in that the individual-level analysis predicts a specific contribution level for each participant. Looking at individual-level predictions, the RPNE correctly predicts half of all choices exactly (chance would predict one sixth). Focussing on those for whom the RPNE pre-conditions are fulfilled most closely, this number increases to two thirds. Again, the ‘mutual-knowledge’ assumption seems to be more critical in our

## 5 SUMMARY AND CONCLUSION

data set.

In addition to the above, the individual-level analysis allows to answer a third question: which criterion do participants use for equilibrium selection in case of multiple RPNEs? In the predictions, I followed Fehr and Schmidt (1999) in assuming participants would use a Pareto-dominance criterion to select the RPNE with the highest payoff sum.

But is Pareto-dominance the criterion participants would use as well? The answer is: partially. While the contribution that corresponds to the Pareto-dominant RPNE is the modal choice, it makes up for only about one third of the choices and 58% of the RPNE-consistent choices. Again, the number is somewhat higher for those participants for whom the pre-conditions are fulfilled: 44%, which are 54% of the RPNE-consistent choices. Those who select other RPNE-consistent choices choose either the ‘most pessimistic’ or the ‘average contribution-sum’ RPNE, in equal parts. In other words, equilibrium selection is an unsolved problem for our participants. Looking at the broader perspective, it may be precisely this missing coordination that sparkles the downward-dynamics we observe in the typical finitely-repeated public-good experiment. Conversely, many public-good situations—in the lab as in real life—could be solved at least partially if players could coordinate on an equilibrium-selection criterion beforehand.

Let me conclude with two remarks. First, on the question of whether there are different types of participants who are ‘Nashy’ to different degrees, or whether there is a single type that happens to be more ‘Nashy’ in some situations, and less so, in others. My understanding is that the heterogeneous-types explanation is the most likely one. This understanding is based on the fact that the categorisation into ‘consolidated-’ or ‘floating-preference’ types, and into ‘mutual-knowledge’ or ‘incomplete-information’ types is based on measurements that are unrelated to the predicted interaction. In particular, the classification is *independent of the interaction partner’s Prefs1-responses* (that participants see when making their choice). On top, auxiliary regressions show that participants’ conditional-contribution types generally are not predictive of their ‘STABILITYBELIEFS’ (which determine the ‘mutual-knowledge’/‘incomplete-information’ classification).<sup>29</sup> Hence, the RPNEs faced by ‘Nashy’ types generally also do not differ from those faced by other (more) ‘non-Nashy’ types.

The second remark concerns why the ‘mutual-knowledge’ assumption seems to be so important. My favoured way to understand the finding goes through best-response behaviour. Four fifth of choices that are ‘approximated best-responses’ turn out to be in line with the RPNE prediction for any of the four behavioural types (compared to about one sixth for choices that are not ‘approximated best-responses’). Yet, ‘mutual-knowledge’ types are far more likely than others to

---

<sup>29</sup>Unless a participant is ‘Unclassifiable’; see Appendix A.



play an ‘approximated best-response’ to their reported beliefs. This finding may look surprising because standard economic theory predicts that participants play a best-response to their beliefs irrespective of where the beliefs come from.

I suggest that the psychology behind the findings is the following: ‘Nashy’ participants believe they ‘understand’ the situation they are facing. Thus, they tend to believe that in such a situation, others’ behaviour is stable and predictable. Thus, they trust their expectations about their opponent’s behaviour and best-respond to these expectations. Best-responses to beliefs that are related to others’ revealed preferences are most likely equilibrium actions. For ‘incomplete-information’ types, this account breaks down right at the start: these people tend not to put faith into their (reported) beliefs, and thus, more often do not best-respond. And hence, ‘mutual knowledge’ predicts equilibrium play.

My account of the mechanism leads to an interesting further hypothesis. If there are two situations, A and B, and most people expect situation A to induce more stable behaviour than situation B, then the Nash-equilibrium will be more predictive of behaviour in situation A, *irrespective* of whether behaviour actually is more stable in situation A or not. However, testing this more general prediction is beyond the scope of the present study and left to future research.

## Technical acknowledgements

I computerised the experiments using z-Tree (Fischbacher, 2007), and recruited participants using ORSEE (Greiner, 2004, PUBLIC PREFERENCES) and hroot (Bock et al., 2014, STANDARD) with Mozilla Firefox. The STANDARD experiment was conducted using z-Tree-unleashed (Duch et al., 2020). I used R (R Development Core Team, 2001, 2012; Ihaka, 1998) in combination with Rkward (Rödiger et al., 2012) and RStudio (RStudio Team, 2015) for the data analysis. R packages Exact (Calhoun, 2015, Boschloo-test), plm (Croissant and Millo, 2008) and lmtest (Zeileis and Hothorn, 2002, both for the regression with cluster-robust standard errors in Appendix A), texreg (Leifeld, 2013, conversion of regression output to  $\LaTeX$ ), and doBy (Højsgaard and Halekoh, 2016, calculating groupwise summary statistics) were of particular value. Most of this was done on a computer running on KDE-based (KDE e.V., 2012) Kubuntu, which required the use of wine for the programming of the experiment. The article was written using Kile.

# Appendix

The Appendix is meant for online publication only.

## Appendix A Additional regressions

	Model 1	Model 2	Model 3
(Intercept)	66.09 (3.08)***	70.39 (5.50)***	58.44 (4.28)***
Perfect conditional cooperator	6.62 (4.85)	4.46 (4.49)	6.17 (4.70)
Imperfect conditional cooperator	-4.07 (5.12)	-3.56 (4.80)	-2.41 (5.17)
Non-monotonic imperfect cond'l cooperator	2.64 (7.98)	2.12 (7.27)	5.18 (7.78)
Triangle cooperator	-3.24 (5.71)	2.17 (6.03)	3.64 (6.45)
Altruist	-21.91 (9.76)*	-15.22 (9.86)	-17.02 (10.73)
Unclassifiable	-18.89 (4.82)***	-14.48 (5.30)**	-12.48 (5.41)*
Same type		10.46 (6.26)	4.09 (6.68)
Consolidated preferences		12.98 (4.14)**	10.21 (4.34)*
Same type × consolidated preferences		-14.45 (8.60)	1.33 (9.54)
Other player is a Perf. cond'l cooperator			3.83 (5.02)
Other player is an Imp. cond'l cooperator			-16.27 (4.93)**
Other player is a Non-mon.imp.cond.coop.			-20.28 (6.30)**
Other player is a Triangle cooperator			-18.16 (5.90)**
Other player is an Altruist			-19.39 (11.76)
Other player is Unclassifiable			-28.66 (4.85)***

\*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$

Table 4: Average STABILITYBELIEFS, regressed on own and other's conditional-contribution type and on whether the player has 'consolidated preferences'. Ordinary least squares model with standard errors clustered on participants. The data sets do not include participants' STABILITYBELIEFS with respect to their SIMPG opponents, because these beliefs were not included in the 'mutual-knowledge'/'incomplete-information' classification. Including these beliefs does not change the conclusions by much, though.

## Appendix B Instructions (translated)<sup>30</sup>

### I General instructions

#### General information

You will now participate in an economic experiment. If you read the following explications thoroughly, you can—depending on your choices—earn money. Therefore, it is very important that you read these explications thoroughly.

The instructions you receive from us are for your personal information only. During the experiment, communication is absolutely prohibited. Non-compliance with this rule will lead to exclusion from the experiment and all payments. If you have questions, please raise your hand. We then answer your question at your cubicle.

In this experiment, you will receive money. The amount you receive depends on your decisions and on the decisions of the other experiment participants. Additionally, you receive a compensation of 2 Euros for completing the ensuing questionnaire.

#### The experiment

The experiment you are participating in today consists of six independent parts. In each of these parts, you will be matched with a different participant. In any case, the participants matched to you will be different people. You will not get to know the identities of the participants you are matched with, neither during nor after the experiment. In the same vein, the participants you are matched with will not get to know your identity.

In some of the parts, there are several participant roles. The role you will take on in actual fact in the different parts will be announced only at the end of the experiment. Therefore, you will make all potentially relevant decisions. Similarly, we will announce only at the very end which of the six parts is relevant for payment. Therefore, you have to determine for all parts what you decide in the according roles. At the end, you will be paid according to the decision you have taken in the relevant role of the randomly-drawn part of the experiment.

Your role and the relevant part are determined by the roll of a die by the participant we have randomly chosen to be the person making the lucky draw at the beginning of the experiment.<sup>31</sup> However, we will announce the realisations

---

<sup>30</sup>The German original is available from the author upon request.

<sup>31</sup>The participant making the lucky draw did not take part in the actual experiment and did not get to know anything about it. The participant was merely asked to roll the die three times, record the results on screen as well as on a sheet of paper (the latter was later put up at the wall in the laboratory), and come to the experimenters' room directly afterwards to collect 8 Euros

of the die rolls only at the end of the experiment. Hence you will know only then which of your decisions will be relevant for your payment.

We describe the individual parts directly on the screen. At each point of the experiment, you only receive the description of the according part. We point out to you that your behaviour from one of the earlier parts will possibly be displayed to other participants in a later part. Further, we would like to inform you that the average payoff to be expected from each of the parts is the same.

---

for the faithful completion of the task.

## II On-screen instructions (translated)

Part 1 out of 7

**Part 1:** In this part there are two types of participants, participant A and participant B. Since only participant A has to make a decision, we explain the experiment from participant A's point of view.

In the following, participant A sees 13 situations, one after the other, in which he has to make a decision. In each situation there are 9 options of distributing money between participant A and participant B. Participant A can choose one of the 9 options in each situation.

**Please note:** Possible payoff distributions are depicted as follows: One line shows the payoff of participant A, the other line shows the payoff of participant B. Which line shows which, is randomly determined and differs between each situation. **So, your own payoffs will sometimes appear in the upper line and sometimes in the bottom line.** By clicking on a distribution option with the left mouse button, you make your choice. After your decision a new screen will appear that you exit with the "next"-button.

At the end of the experiment, the roll of a die will determine which of the 13 situations will be paid out and whether you receive your payment in the role of participant A or participant B.

Situation Number : 1 out of 13

The other person receives

15	16	18	19	20	21	23	24	25
23	21	19	18	16	15	13	12	10

The other person receives

Part 1: Screenshot of the instruction stage.

Part 2 out of 7

**Part 2:** In this part of the experiment you form a group of 2 with another randomly assigned participant. Each participant has an endowment of 15 Euros which he can either transfer to his private account or invest in a project. The whole 15 Euros or a fraction of it can be invested in the project. Money that is not invested in the project will automatically be transferred to the private account.

Every Euro invested in the project by a group member yields 1.33 Euros for the group. *Both group members (including the one investing in the project) receive 0.67 Euros of that amount, irrespective of whether and how much a group member invested in the project.* For every Euro that a participant transfers to his private account, he receives 1 Euro. The payoff of the other group member is not affected by that.

Therefore, your payoff is calculated as follows:

**Your payoff =**  
 money transferred to the private account  
 + 0.67 \* sum of the amounts that you and the other group member invested in the project

This results in the following **situation features:**

- The more is invested in the group project, the bigger is the total sum of payoffs of the two group members.
- However, for every possible investment of the other group member your personal payoff is higher the less you invest in the group project.

**Please familiarize yourself with the situation now, by entering in the two boxes different investments you and the other group member might make. Afterwards, click on the button "calculate". For practical reasons, you can only invest 0, 3, 6, 9, 12 or 15 Euros in the project.**

Your possible investment:

Possible investment of the other group member:

Your investment	Your payoff	Investment of the other	Payoff of the other

Part 2, PREFS1: instructions; screen with only text (as in upper half) omitted.

Now please answer the following questions:

In case you need a calculator, please click on the corresponding icon below.

1. Assume that no group member invests anything into the project.  
 How many Euros does each participant receive?

Part 2, PREFS1: instructions; comprehension question 1 (upper part as above).

Now please answer the following questions:

In case you need a calculator, please click on the corresponding icon below.

2. Assume that one group member invests the whole 15 Euros in the project. The other group member doesn't invest anything.

How many Euros does the investing group member receive?

How many Euros does the non-investing group member receive?

Part 2, PREFS1: instructions; comprehension question 2 (upper part as above).

Now please answer the following questions:

In case you need a calculator, please click on the corresponding icon below.

3. Assume that both group members invest the whole 15 Euros in the project.  
 How many Euros does each participant receive?

Part 2, PREFS1: instructions; comprehension question 3 (upper part as above).

Part 2 out of 7

Like in part 1, there are again two types of participants, A and B, and again you don't know which type you are. One group member will be of type A, the other group member will be of type B. First, participant A makes his investment decision. In a second step, participant B makes his investment decision. However, he can adjust his decision depending on participant A's decision.  
For practical reasons, you can only invest 0, 3, 6, 9, 12 or 15 Euros in the project.

**You are now taking your decisions as participant A.**

Amount that is available to you: 15  
Your investment in the project:

Part 2, Prefs1: unconditional-contribution choice.

Part 2 out of 7

**You are now taking your decisions as participant B.**

In the following table, the rows mark the potential decisions of participant A. Please state how you react to each individual investment decision by participant A (marked in yellow in the following).

**Please click on how many Euros you invest into the project. Note: by clicking on an investment, you already take your decision.**

**Investment by A:**

0	For you: 15	For you: 14	For you: 13	For you: 12	For you: 11	For you: 10
	For A: 15	For A: 17	For A: 19	For A: 21	For A: 23	For A: 25
3	For you: 17	For you: 16	For you: 15	For you: 14	For you: 13	For you: 12
	For A: 14	For A: 16	For A: 18	For A: 20	For A: 22	For A: 24
6	For you: 19	For you: 18	For you: 17	For you: 16	For you: 15	For you: 14
	For A: 13	For A: 15	For A: 17	For A: 19	For A: 21	For A: 23
9	For you: 21	For you: 20	For you: 19	For you: 18	For you: 17	For you: 16
	For A: 12	For A: 14	For A: 16	For A: 18	For A: 20	For A: 22
12	For you: 23	For you: 22	For you: 21	For you: 20	For you: 19	For you: 18
	For A: 11	For A: 13	For A: 15	For A: 17	For A: 19	For A: 21
15	For you: 25	For you: 24	For you: 23	For you: 22	For you: 21	For you: 20
	For A: 10	For A: 12	For A: 14	For A: 16	For A: 18	For A: 20

**Your investment:**

Part 2, Prefs1: conditional-contribution choice (preference elicitation).

Part 3 out of 7

**Part 3:** Like in the previous parts, you will interact with a new randomly assigned participant also in the 3rd part. The situation is the same as in the 2nd part. Therefore, you can again invest 0, 3, 6, 9, 12 or 15 Euros in a project. Your payoff is the result of:

**Your payoff =**  
**money transferred to the private account**  
**+ 0.67\*sum of the amounts that you and the other group member invested in the project**

This results in the following **situational features**:

The more is invested in the group project, the bigger is the total sum of payoffs of the two group members.

However, for every possible investment of the other group member your personal payoff is higher the less you invest in the group project.

As before, participant A makes his investment decision first. And again, participant B can adjust his decision depending on participant A's decision.

[next](#)

[I need help/have a question](#)

Part 3, Prefs2: instructions.

Part 3 out of 7

**Part 3:** Like in the previous parts, you will interact with a new randomly assigned participant also in the 3rd part. The situation is the same as in the 2nd part. Therefore, you can again invest 0, 3, 6, 9, 12 or 15 Euros in a project. Your payoff is the result of:

**Your payoff =**  
**money transferred to the private account**  
**+ 0.67\*sum of the amounts that you and the other group member invested in the project**

**You are now taking your decisions as participant A.**

Amount that is available to you: 15

Your investment in the project:

[OK](#)

[I need help/have a question](#)

Part 3, Prefs2: unconditional-contribution choice.



Part 3 out of 7

You are now taking your decisions as participant B.

**Investment by A:**

0	For you: 15	For you: 14	For you: 13	For you: 12	For you: 11	For you: 10
	For A: 15	For A: 17	For A: 19	For A: 21	For A: 23	For A: 25
3	For you: 17	For you: 16	For you: 15	For you: 14	For you: 13	For you: 12
	For A: 14	For A: 16	For A: 18	For A: 20	For A: 22	For A: 24
6	For you: 19	For you: 18	For you: 17	For you: 16	For you: 15	For you: 14
	For A: 13	For A: 15	For A: 17	For A: 19	For A: 21	For A: 23
9	For you: 21	For you: 20	For you: 19	For you: 18	For you: 17	For you: 16
	For A: 12	For A: 14	For A: 16	For A: 18	For A: 20	For A: 22
12	For you: 23	For you: 22	For you: 21	For you: 20	For you: 19	For you: 18
	For A: 11	For A: 13	For A: 15	For A: 17	For A: 19	For A: 21
15	For you: 25	For you: 24	For you: 23	For you: 22	For you: 21	For you: 20
	For A: 10	For A: 12	For A: 14	For A: 16	For A: 18	For A: 20

In the following table, the rows again mark the potential decisions of participant A. Please state how you react to each individual investment decision by participant A (marked in yellow in the following).

**Please click on how many Euros you invest into the project. Note: by clicking on an investment, you already take your decision.**

**Your investment:**

Part 3, PREFS2: conditional-contribution choice (preference elicitation).

Part 3 out of 7

**You just stated how you react to every possible investment decision by participant A. Now we want you to assess how participant A behaved. In the following, you have to state the probability with which participant A chose each investment decision.**

In part 3, participant A had to choose between 6 different investment levels. So, for each of these investment levels you have to state the probability with which participant A chose this exact level. Doing so, you can earn an additional 2.00 Euros. Your chance of earning 2.00 Euros increases with the accuracy of your assessment. Your assessment is the better, the closer it matches the actual behavior of participant A. By clicking on the button "details", you can see how we calculate the payoff that rewards the accuracy of your assessment. If you are not interested in the details, you can safely ignore the corresponding explanation.

**What is important for you is just that the chance of getting a high payoff is maximized when you correctly assess the choice behavior of participant A. We want you to have an incentive to properly think about the behavior of participant A and to be rewarded if you capture it well and state it accordingly.**

You can state your assessment of the probabilities with the help of a bar chart. Below you see an example of such a chart. If you click on the chart, you will notice that you can adjust the height of the bars by clicking on them. At the same time, in the upper right of the diagram a window will open up with which you can define the exact height of the bar. The sum of the bars of the probability distribution always has to equal 100 percent. If you click on "scale", the bars adjust accordingly, while the relative size ratio remains the same. If, by using exact entries, the sum is already 100 percent, the bars are already correctly scaled. Please note that the percent values above the bars are rounded to integers, so that the sum can differ slightly from 100.

**How likely is it that participant A chose an investment of X?**

**Example chart:**

When you click on "next", such a diagram will appear on your screen. Please note that this diagram can be relevant for your payoff. Please work on the diagram on the next page. You can return to these instructions at any time. Afterwards, the 3rd part of the experiment is completed.

Part 3, PREFS2: instructions for the belief-elicitation on A's choice (training for the SIMPGBELIEFS and STABILITYBELIEFS experiments).

Part 3 out of 7

**Payoff of the probability assessment:**

In case this part will be relevant for your payoff, you will receive an additional payoff of 2 Euros with a certain probability  $p$  and no additional payoff with probability  $1-p$ . Your assessments affect your chance  $p$  in the following way:

First, your chance  $p$  is set to 100%. Then there will be deductions for wrong assessments. The deviations are first divided by 100, then squared, then halved and then subtracted from 1. The result is multiplied by 100 to get a percentage between 0 and 100.

If you, for example, assigned 70% to a certain investment, but participant A chose a different investment level, your deviation is 0.7. So  $0.70 * 0.70 / 2 = 0.245$  (approx. 25%) is subtracted from  $p$ .

For the investment level that participant A actually chose, it is bad if your assessment is far away from 100%. That deviation will also be squared, halved and subtracted. If you, for example, assigned only 60% to the correct investment,  $(1 - 0.60) * (1 - 0.60) / 2 = 0.40 * 0.40 / 2 = 0.08$  (8%) will be subtracted from  $p$ .

In the end, all the subtractions are added up. The smaller these squared deviations are, the better was your assessment. For those who are interested: Here's the mathematical formula that is used to calculate the quality of your assessment and thereby the chance  $p$  to earn 2.00 Euros:

$$p = (1 - 0.5 * (\text{sum of the squared probabilities on investment levels that were not chosen by participant A} + \text{squared counter-probability of the actually chosen investment level})) * 100$$

The computer program will calculate the quality of your assessment  $p$  and show it to you. Your assessment was the better and the chance to earn additional 2.00 Euros in this part is the higher, the higher your value of  $p$  is. If part 3 is relevant for your payoff, a number between 0 and 100 will be determined by rolling dice at the end of the experiment. If this number is smaller or equal to  $p$ , you receive 2.00 additional Euros. If the number is larger than  $p$ , you won't receive an additional payoff.

**Summary:**

**To have a good chance of receiving the additional payoff, it should be your goal to get as few subtractions from  $p$  as possible. The best way to achieve that is to make a good assessment of the choice behavior of participant A and to state it truthfully.**

[Hide details](#)

[I need help/have a question](#)

Part 3, PREFS2: instructions for the belief-elicitation on A's choice, details (training for the SIMPGBELIEFS and STABILITYBELIEFS experiments).

**You just stated how you react to every possible investment decision by participant A. Now we want you to assess how participant A behaved. In the following, you have to state the probability with which participant A chose each investment decision.**

In part 3 participant A had to choose between 6 different investment levels. So, for each of these investment levels you have to state the probability with which participant A chose this exact level. Doing so, you can earn additional 2.00 Euros. Your chance of earning 2.00 Euros increases with the accuracy of your assessment. Your assessment is all the better, the more it matches the actual behavior of participant A. By clicking on the button "details" you can see the calculation of the payoff that rewards the accuracy of your assessment. If you are not interested in the details, you can ignore the corresponding explanation.

**What is important for you is just that the chance of getting a high payoff is maximized when you correctly assess the decision behavior of participant A. We want you to have an incentive to properly think about the behavior of participant A and to be rewarded if you capture it well and state it accordingly.**

You can state your assessment of the probabilities with the help of a bar chart. Below you see an example of such a chart. If you click on the chart, you will notice that you can adjust the height of the bars by clicking on them. At the same time in the upper right of the diagram a window will open up with which you can define the exact height of the bar. The sum of the bars of the probability distribution always has to equal 100 percent. If you click on "scale", the bars adjust accordingly, while the relative size ratio stays the same. If, by using exact entries, the sum is already 100 percent, the bars are already correctly scaled. Please note that the percent values above the bars are rounded to integers, so the sum can slightly differ from 100.

[back to instructions](#)

How likely is it that participant A chose an investment of X?



[next](#)

Part 3, PREFS2: belief-elicitation on A's choice (training for the SIMPGBELIEFS and STABILITYBELIEFS experiments).

Part 4 out of 7

**Reaction of your newly assigned group member as B in part 2:**

Investment of A:	0	3	6	9	12	15
0	For B: 15 For A: 15	For B: 14 For A: 17	For B: 13 For A: 19	For B: 12 For A: 21	For B: 11 For A: 23	For B: 10 For A: 25
3	For B: 17 For A: 14	For B: 16 For A: 16	For B: 15 For A: 18	For B: 14 For A: 20	For B: 13 For A: 22	For B: 12 For A: 24
6	For B: 19 For A: 13	For B: 18 For A: 15	For B: 17 For A: 17	For B: 16 For A: 19	For B: 15 For A: 21	For B: 14 For A: 23
9	For B: 21 For A: 12	For B: 20 For A: 14	For B: 19 For A: 16	For B: 18 For A: 18	For B: 17 For A: 20	For B: 16 For A: 22
12	For B: 23 For A: 11	For B: 22 For A: 13	For B: 21 For A: 15	For B: 20 For A: 17	For B: 19 For A: 19	For B: 18 For A: 21
15	For B: 25 For A: 10	For B: 24 For A: 12	For B: 23 For A: 14	For B: 22 For A: 16	For B: 21 For A: 18	For B: 20 For A: 20

**Part 4:** In this part of the experiment you once more form a group of 2 with another randomly assigned participant. The situation is the same as in part 2 and 3. However, both group members now make their decisions simultaneously, so that there are no different types of participants. Furthermore, you get information about how your newly assigned group member behaved in part 2 as participant B, before you make your decision. You will find this information in the table following on the right.

Complete situation description part 2

**You now make your decision.**

Amount that is available to you: 15

Your investment in the project:

OK

I need help/have a question

Part 4, SIMPG: contribution choice; screen with only instructions (as in upper half) omitted.

Part 5 out of 7

**Reaction of the other participant as B in part 2:**

Investment of A:	0	3	6	9	12	15
0	For B: 15 For A: 15	For B: 14 For A: 17	For B: 13 For A: 19	For B: 12 For A: 21	For B: 11 For A: 23	For B: 10 For A: 25
3	For B: 17 For A: 14	For B: 16 For A: 16	For B: 15 For A: 18	For B: 14 For A: 20	For B: 13 For A: 22	For B: 12 For A: 24
6	For B: 19 For A: 13	For B: 18 For A: 15	For B: 17 For A: 17	For B: 16 For A: 19	For B: 15 For A: 21	For B: 14 For A: 23
9	For B: 21 For A: 12	For B: 20 For A: 14	For B: 19 For A: 16	For B: 18 For A: 18	For B: 17 For A: 20	For B: 16 For A: 22
12	For B: 23 For A: 11	For B: 22 For A: 13	For B: 21 For A: 15	For B: 20 For A: 17	For B: 19 For A: 19	For B: 18 For A: 21
15	For B: 25 For A: 10	For B: 24 For A: 12	For B: 23 For A: 14	For B: 22 For A: 16	For B: 21 For A: 18	For B: 20 For A: 20

**Part 5:** With which probability did the other participant choose each investment decision in part 4? For assistance, you can see the participant-B-decisions from the 2nd part made by the participant that was randomly assigned to you in the 4th part.

Like in part 3, you can earn money with your assessment of the behavior of the other participant. You receive either 20.00 or 4.00 Euros. The chance to earn 20.00 Euros increases with the accuracy of your assessment. The chance of earning 20.00 Euros is calculated like the chance of earning the extra amount of 2.00 Euros in part 3. You can see the details of the calculation of your chance to earn 20.00 Euros by clicking on "details". If you are not interested in the details, you can ignore the corresponding explanation.

**What is important for you to know is that the chance of getting a high payoff is maximized when you correctly assess the decision behavior of the other participant. We want you to have an incentive to properly think about the behavior of the other participant and to be rewarded if you capture it well and state it accordingly.**

details

**How likely is it that the other participant chose an investment of X in part 4?**

Percentage sum: 0

scale

next

I need help/have a question

Part 5, SIMPGBELIEFS: belief elicitation; screen with only instructions (as in upper half) omitted. Clicking on "details" led to an analogous screen as in Part 3.

Part 6 out of 7

**Part 6:** In part 2 and part 3 you and the other participants had to make an investment decision, where you, as participant B, could adjust your decision depending on the decisions of your randomly assigned participant A. In part 6 we want you to assess the behavior of four other type-B-participants in part 3. In the following, you have to state with which probability these participants made their investment decisions, depending on the behavior of the corresponding type-A-participants. For assistance, you can see the behavior of each participant in the role of participant B in part 2.

Each of the four B-participants had to state how they would react to each of the 6 possible investment decisions of the corresponding A-participant. So, for each of the 6 possible investment decisions you have to make an assessment of the probability with which the respective B-participant chose a certain reaction. Therefore, this time you see 6 diagrams for each of the four B-participants.

Like in the previous parts, you can earn money with your assessment of the behavior of the other participants. For each assessment you receive either 6.00 Euros or nothing. The chance to earn 6.00 Euros increases with the accuracy of your assessment. First, for each of the four B-participants one of the 6 diagrams is randomly drawn. The chance of earning 6.00 Euros for that randomly drawn diagram of the respective B-participant is then calculated in the same way the chance of earning the extra amount of 20.00 Euros in part 5 was calculated. Again, you can see the details of the calculation of your chance to earn 6.00 Euros each, by clicking on "details". If you are not interested in the details, you can ignore the corresponding explanation.

**What is important for you to know is that the chance of getting a high payoff is maximized when you correctly assess the decision behavior of the other participants. We want you to have an incentive to properly think about the behavior of the B-participants and to be rewarded if you capture it well and state it accordingly.**

details

next

I need help/have a question

Part 6, STABILITYBELIEFS: instructions; "details" led to an analogous screen as in Part 3.

Part 6 out of 7

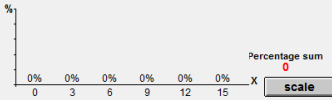
On the right, you see the behavior as participant B in part 2 of the first participant for whom you have to make a guess on the behavior in part 3.

Instructions

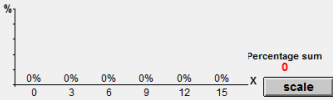
Reaction of the respective participant as B in part 2:

Investment of A:	0						3						6						9						12						15					
	For B:		For A:		For B:		For A:		For B:		For A:		For B:		For A:		For B:		For A:		For B:		For A:		For B:		For A:		For B:		For A:					
0	15	15	14	17	13	19	12	21	11	23	10	25	17	16	16	16	15	14	14	13	13	12	12	11	11	10	10	9	9	8						
3	17	14	16	16	15	18	14	20	13	22	12	24	19	18	17	17	16	16	15	15	14	14	13	13	12	12	11	11	10	10						
6	19	13	18	15	17	17	16	19	15	21	14	23	21	12	20	14	19	18	18	17	17	16	16	15	15	14	14	13	13	12						
9	21	12	20	14	19	17	18	18	17	20	16	22	12	11	11	13	10	13	9	12	8	11	7	10	6	9	5	8	4	7						
12	23	11	22	13	21	15	20	17	19	19	18	21	11	25	10	24	9	23	8	22	7	21	6	20	5	19	4	18	3	17						
15	25	10	24	12	23	14	22	16	21	18	20	20	19	21	18	22	17	21	16	20	15	19	14	18	13	17	12	16	11	15						

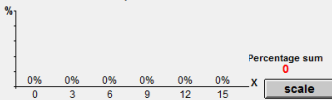
How likely is it that the participant's reaction on an investment of 0 was X in part 3?



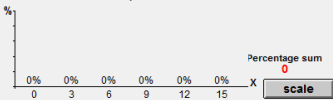
How likely is it that the participant's reaction on an investment of 3 was X in part 3?



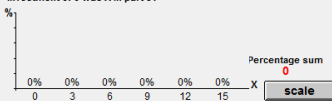
How likely is it that the participant's reaction on an investment of 6 was X in part 3?



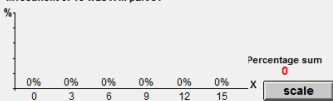
How likely is it that the participant's reaction on an investment of 9 was X in part 3?



How likely is it that the participant's reaction on an investment of 12 was X in part 3?



How likely is it that the participant's reaction on an investment of 15 was X in part 3?



next

I need help/have a question

Part 6, STABILITYBELIEFS: belief elicitation.

Part 7 out of 7

**Part 7:** In this part of the experiment you once more form a group of 2 with another randomly assigned participant. The situation is a mixture of parts 2, 3 and 4. Participant A and participant B make their decisions successively, whereby participant B can make his decision dependent on participant A's decision. Before you make your decision as participant A, you learn about how the participant that was randomly assigned to you behaved as participant B in part 2. You will find this information in a table on the right. Participant B, however, doesn't learn about your behavior in part 2.

*Keep in mind:* The participants make their decisions successively, so participant B can make his decision depending on the decision of participant A.

**Complete situation description**

**Reaction of the newly assigned participant as B in part 3:**

	0	3	6	9	12	15
0	For B: 15 For A: 15	For B: 14 For A: 17	For B: 13 For A: 19	For B: 12 For A: 21	For B: 11 For A: 23	For B: 10 For A: 25
3	For B: 17 For A: 14	For B: 16 For A: 16	For B: 15 For A: 18	For B: 14 For A: 20	For B: 13 For A: 22	For B: 12 For A: 24
6	For B: 19 For A: 13	For B: 18 For A: 15	For B: 17 For A: 17	For B: 16 For A: 19	For B: 15 For A: 21	For B: 14 For A: 23
9	For B: 21 For A: 12	For B: 20 For A: 14	For B: 19 For A: 16	For B: 18 For A: 18	For B: 17 For A: 20	For B: 16 For A: 22
12	For B: 23 For A: 11	For B: 22 For A: 13	For B: 21 For A: 15	For B: 20 For A: 17	For B: 19 For A: 19	For B: 18 For A: 21
15	For B: 25 For A: 10	For B: 24 For A: 12	For B: 23 For A: 14	For B: 22 For A: 16	For B: 21 For A: 18	For B: 20 For A: 20

You are now making your decision as participant A.

Amount that is available to you: 15

Your investment in the project:

**OK**

I need help/have a question

Part 7, PREFS3: unconditional contribution; "Complete situation description" led to a screen similar to the instructions screen in Part 3.

Part 7 out of 7

You are now making your decisions as participant B.

Have in mind that these decisions could be relevant for your payoff!

**Investment of A:**

0	For you: 15 For A: 15	For you: 14 For A: 17	For you: 13 For A: 19	For you: 12 For A: 21	For you: 11 For A: 23	For you: 10 For A: 25
3	For you: 17 For A: 14	For you: 16 For A: 16	For you: 15 For A: 18	For you: 14 For A: 20	For you: 13 For A: 22	For you: 12 For A: 24
6	For you: 19 For A: 13	For you: 18 For A: 15	For you: 17 For A: 17	For you: 16 For A: 19	For you: 15 For A: 21	For you: 14 For A: 23
9	For you: 21 For A: 12	For you: 20 For A: 14	For you: 19 For A: 16	For you: 18 For A: 18	For you: 17 For A: 20	For you: 16 For A: 22
12	For you: 23 For A: 11	For you: 22 For A: 13	For you: 21 For A: 15	For you: 20 For A: 17	For you: 19 For A: 19	For you: 18 For A: 21
15	For you: 25 For A: 10	For you: 24 For A: 12	For you: 23 For A: 14	For you: 22 For A: 16	For you: 21 For A: 18	For you: 20 For A: 20

In the following table, the rows mark the potential decisions of participant A. Please state how you react to each individual investment decision by participant A (marked in yellow in the following).

**Now please click on how many Euros you invest into the project.**

Your investment:

I need help/have a question

Part 7, PREFS3: conditional contribution (preference elicitation).

## References

- Ambrus, A. and Pathak, P. A. (2011). Cooperation over finite horizons: A theory and experiments. *Journal of Public Economics*, 95(7-8):500–512.
- Andreoni, J. and Samuelson, L. (2006). Building rational cooperation. *Journal of Economic Theory*, 127:117–154.
- Arifovic, J. and Ledyard, J. O. (2012). Individual evolutionary learning, contributions mechanism: other-regarding preferences, and the voluntary contributions mechanism. *Journal of Public Economics*, 96:808–823.
- Aumann, R. and Brandenburger, A. (1995). Epistemic conditions for nash equilibrium. *Econometrica*, 5:1161–1180.
- Blanco, M., Engelmann, D., and Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, 72(2):321 – 338.
- Bock, O., Baetge, I., and Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, 71:117–120.
- Brunner, C., Kauffeldt, T. F., and Rau, H. (2021). Does mutual knowledge of preferences lead to more nash equilibrium play? experimental evidence. *European Economic Review*, 135:103735.
- Calhoun, P. (2015). *Exact: Unconditional Exact Test*. R package version 1.6.
- Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117(3):p817 – 869.
- Croissant, Y. and Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, 27(2).
- Cubitt, R. P., Starmer, C., and Sugden, R. (2001). Discovered preferences and the experimental evidence of violations of expected utility theory. *Journal of Economic Methodology*, 8(3):385–414.
- Drouvelis, M., Metcalfe, R., and Powdthavee, N. (2015). Can priming cooperation increase public good contributions? *Theory and Decision*, 79(3):479–492.
- Duch, M. L., Grossmann, M. R., and Lauer, T. (2020). z-tree unleashed: A novel client-integrating architecture for conducting z-tree experiments over the internet. *Journal of Behavioral and Experimental Finance*, 28:100400.
- Dufwenberg, M., Gächter, S., and Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior*, 73(2):459–478.
- Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3):817–868.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

- Fischbacher, U. and Gächter, S. (2010). Social preferences, beliefs and the dynamics of free riding in public good experiments. *American Economic Review*, 100(1):541–556.
- Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3):397–404.
- Fischbacher, U., Gächter, S., and Quercia, S. (2012). The behavioral validity of the strategy method in public good experiments. *Journal of Economic Psychology*, 33(4):897 – 913.
- Greiner, B. (2004). An online recruitment system for economic experiments. In Kremer, K. and Macho, V., editors, *Forschung und wissenschaftliches Rechnen 2003*, volume 63, pages 79–93. Ges. für Wiss. Datenverarbeitung, Göttingen.
- Guala, F., Mittone, L., and Ploner, M. (2013). Group membership, team preferences, and expectations. *Journal of Economic Behavior & Organization*, 86:183–190.
- Healy, P. J. (2011). Epistemic foundations for the failure of Nash equilibrium. Working paper.
- Højsgaard, S. and Halekoh, U. (2016). *doBy: Groupwise Statistics, LSmeans, Linear Contrasts, Utilities*. R package version 4.5-15.
- Hossain, T. and Okui, R. (2013). The binarized scoring rule. *Review of Economic Studies*, 80(3):984–1001.
- Ihaka, R. (1998). R: Past and future history. In Weisberg, S., editor, *Proceedings of the 30th Symposium on the Interface*, pages 392–396.
- Kamei, K. (2016). Democracy and resilient pro-social behavioral change: an experimental study. *Social Choice and Welfare*, 47:359–378.
- Kartal, M. and Müller, W. (2021). A new approach to the analysis of cooperation under the shadow of the future: Theory and experimental evidence. Working paper.
- KDE e.V. (2012). About kde. <http://www.kde.org/community/whatiskde/>. Last accessed on 31st January, 2013.
- Leifeld, P. (2013). texreg: Conversion of statistical model output in R to  $\LaTeX$  and HTML tables. *Journal of Statistical Software*, 55(8):1–24.
- Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, 1(3):593–622.
- McKelvey, R. D. and Page, T. (1990). Public and private information: An experimental study of information pooling. *Econometrica*, 58(6):1321–1339.
- Murphy, R. O., Ackermann, K. A., and Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6(8):771–781.

- Offerman, T., Sonnemans, J., and Schram, A. (1996). Value orientations, expectations and voluntary contributions in public goods. *The Economic Journal*, 106(437):817–845.
- R Development Core Team (2001). What is R? *R News*, 1:2–3.
- R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- Rödiger, S., Friedrichsmeier, T., Kapat, P., and Michalke, M. (2012). Rkward: A comprehensive graphical user interface and integrated development environment for statistical analysis with r. *Journal of Statistical Software*, 49(9):1–34.
- RStudio Team (2015). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.
- Spadaro, G., Tiddi, I., Columbus, S., Jin, S., Ten Teije, A., and Balliet, D. (2020). The cooperation databank. Psyarxiv.
- Vespa, E., Weidman, T., and Wilson, A. (2021). Testing models of strategic uncertainty: Equilibrium selection in repeated games. Working paper.
- Wolff, I. (2015). When best-replies are not in equilibrium: Understanding cooperative behaviour. Research Paper Series 97, Thurgau Institute of Economics.
- Wolff, I. (2017). What are the equilibria in public-good experiments? *Economics Letters*, 150:83–85.
- Zeileis, A. and Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, 2(3):7–10.