Received: 19 March 2024 | Accepted: 15 January 2025

DOI: 10.1111/ecin.13280

ORIGINAL ARTICLE

Economic Inquiry

Predicting voluntary contributions by "revealed-preference Nash-equilibrium"

Irenaeus Wolff

Thurgau Institute of Economics (TWI) / University of Konstanz, Kreuzlingen, Switzerland

Correspondence
Irenaeus Wolff.
Email: wolff@twi-kreuzlingen.ch

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 100018_152788

Abstract

One-shot public-good situations are prominent in the public debate, and a prime example for behavior diverging from the standard Nash-equilibrium. Could a Nash-equilibrium predict one-shot public-good behavior in principle? A "revealed-preference Nash-equilibrium" (RPNE) out-of-sample predicts behavior, 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. While the Pareto-dominant equilibrium is the modal choice, many participants use other criteria. Given the predictive positive-contributions RPNEs, many real-life public-good problems may be solvable if players could coordinate on an equilibrium-selection criterion beforehand.

KEYWORDS
best-response, conditional cooperation, knowledge of preferences, Nash-equilibrium, preference stability, public good, social dilemma, social preferences

JEL CLASSIFICATION
C72, C92, D83, H41

1 | INTRODUCTION

One-shot public-good situations are extremely prominent in both economic textbooks and popular non-scientific conceptualizations of some of the most pressing problems humanity is facing (e.g., when it comes to efforts to contain climate change). At the same time, one-shot public-good experiments are a prime example for a situation in which people's behavior 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

Abbreviations: AL (2012), Arifovic and Ledyard (2012); FS (1999), Fehr and Schmidt (1999); MPCR, marginal per-capita return; *N*, number of observations; RPNE, revealed-preference Nash-equilibrium; RQ, research question; USD, US-dollars.

Managing Editor: Stefano Barbieri

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; see, e.g., Rapoport, 1985 for a similar argument).[1]

Even so, it remains unclear whether the difference between behavior and the standard Nash-equilibrium is due to a misspecification of the players' preferences (who, for example, may take others' payoffs into consideration), a mistaken account of the strategic aspects of the interaction, or both (see, for example, the analysis in Rapoport & Suleiman, 1992). As a consequence, a good account of behavior 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 behavior successfully 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 preconditions 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 behavior (Brunner et al., 2021; Healy, 2011).

In the context of the question raised above, in two-player games it is not the strategic aspect that makes behavior diverge from the standard Nash-equilibrium (this changes to some degree when we move to three-player games). As I will show, the missing knowledge of others' preferences does play a role, but the main "culprit" is the specification of players' preferences. However, capturing participants' preferences is by no means trivial, as a comparison to the models by Fehr and Schmidt (1999) and, in particular, by Arifovic and Ledyard (2012) shows.

The finding that positive-contributions equilibria are meaningful for behavior opens new questions. 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 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?

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 behavior 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—could coordinate on an equilibrium-selection criterion beforehand, many public-good problems potentially could be solved.

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. More recently, Gächter et al. (2017) called the approach the "ABC of cooperation": "attitudes" (a_i; conditional-contribution preferences) together with "beliefs" (b_i) determine "effective contributions" (in a simultaneous public-good situation, c_i), so that cooperation is explained as $a_i(b_i) \rightarrow c_i$. All three studies clearly establish the behavioral validity of conditional-contribution preference measurements for actual public-good play.

In addition, Fischbacher and Gächter (2010) show two aspects that are important to understand the contribution of the present paper. First, beliefs are predictive for public-good contributions *on top of* the "predicted contributions" $a_i(b_i)$. In this sense, beliefs enter contributions twice. And second, players' beliefs are not equilibrium beliefs (and participants update the beliefs suboptimally). To restate, beliefs are particularly important for predicting public-good contributions, and those important beliefs are non-equilibrium beliefs. What the RPNE does is to take exactly those beliefs out of the equation and to substitute them by the standard equilibrium assumptions. In the above framework, cooperation becomes simply $\mathbf{a} \rightarrow c_i$.

In a similar vein, 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 situations 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 × 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 behavior 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 × 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 behavior.

In the end, the paper's contribution to the literature is multi-faceted: it provides a proof of concept for the "revealed-preference Nash equilibrium" introduced in Wolff (2017), showing that it outperforms other popular models in out-of-sample predictions particularly in two-player games. By that, it shows that an equilibrium prediction can be successful even in one-shot social-dilemma situations and even when players do not know their opponents' preferences. Given that a substantial number of those equilibria have positive contributions, it opens up the question of why contributions typically decline toward zero in repeated settings. And it shows that there is substantial heterogeneity in the equilibrium-selection criteria people use, which hints at an answer to the above question.

2 | REVEALED-PREFERENCE NASH EQUILIBRIUM

It is well-known that social preferences play a role for behavior, 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 social-preference utility functions (e.g., Levine, 1998; Fehr & 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).

Definition 1. Let $\mathcal{A}_i$ denote the (finite) set of agent i's actions in an n-player game and let $\mathcal{A}_{-i} \quad \mathcal{A}_1 \times \mathcal{A}_2 \times \cdots \times \mathcal{A}_{i-1} \times \mathcal{A}_{i+1} \times \cdots \times \mathcal{A}_{n-1} \times \mathcal{A}_n$ denote the set of all action profiles of all agents but i. Furthermore, let $r_i : \mathcal{A}_{-i} \rightarrow \mathcal{A}_i$ denote agent i's observed reaction to all possible action profiles of the other agents. Then, an RPNE of the game is an action profile $(a_1, a_2, ..., a_{n-1}, a_n)$ that satisfies $a_i \quad r_i((a_1, ..., a_{i-1}, a_{i+1}, ..., a_n)), \forall i$.[2]

The RPNE is the set of Nash-equilibria that results when measured conditional-contribution preferences are interpreted as best-reply correspondences. More precisely, an RPNE of the simultaneous game is a contribution profile in which each player chooses a contribution in line with her conditional-contribution preferences, given the other players' contributions (which are themselves conditional on the contributions in the group). Note that by interpreting the elicited conditional-contribution vectors as best-reply correspondences directly (as first suggested by Rapoport & Suleiman, 1992), I bypass the question of how to model the preferences in an expected-utility framework. One potential way of doing so would be to use a model like the one of Falk and Fischbacher (2006).

The intuition behind the RPNE 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 $\mathbf{x}$.[3]

To make things very clear, let me provide an example. Imagine there is a (tiny) population of 4 people who participate in a two-player variant of the conditional-contributions-elicitation experiment of Fischbacher et al. (2001) with three contribution levels $\{0,1,2\}$ and a given marginal per-capita return (MPCR). Two of them are perfect conditional-cooperators and thus react to each unconditional contribution $c_j^{(u)}$ by the other player by $c_i^{(c)}(c_j^{(u)}) \quad c_j^{(u)}$, one person is a complete free-rider who always chooses $c_i^{(c)}(c_j^{(u)}) \quad 0$, and the fourth person an imperfect conditional cooperator who chooses $c_i^{(c)}(0) \quad 0$, $c_i^{(c)}(1) \quad 1$, and $c_i^{(c)}(2) \quad 1$. Equating $c_i^{(c)}(c_j^{(u)}) \quad r_i(c_j)$, that is, with the response function to the opponent's action c_j in a simultaneous game, and invoking Pareto-dominance, we would expect two perfect conditional contributors to choose $\mathbf{c} \quad (2,2)$, a perfect conditional contributor and a free-rider to choose $\mathbf{c} \quad (0,0)$, a perfect conditional contributor and an imperfect conditional contributor (of the observed variant) to choose $\mathbf{c} \quad (1,1)$, and an imperfect conditional contributor and a free-rider to choose $\mathbf{c} \quad (0,0)$.

Given that the proportions of free-riders, perfect and imperfect conditional contributors *in our example population* are 25%, 50%, and 25%, respectively, we obtain the population-level prediction for the simultaneous game given in Table 1: $\{0,1,2\}$ would be chosen in 43.75%, 31.25%, and 25% of the cases, respectively. Figure 1 shows the predictions *for the actual parameters* for which the RPNE has been calculated (note that the predictions are likely to change—and do change—for different parameters; in particular, for an MPCR close to 0, we would not expect to see hardly any conditional cooperators, while for an MPCR close to 1, there would be many).

The predictions shown in Figure 1 were calculated already in Wolff (2017).[4] The first contribution of the present paper is to provide the first test of RPNE predictions on actual data from existing public-good experiments as well as new experiments, and to compare that to relevant other models providing quantitative predictions. In particular, I show that the RPNEs calculated in Wolff (2017) are predictive for behavior 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., 2010; Drouvelis et al., 2015; Dufwenberg et al., 2011, for three-player games). I focus on predictions because only predictive success accross different situations shows whether a model robustly identifies a relevant mechanism (as opposed to being able to fit data due to a flexibility that rests on free parameters).

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 behavior well. Thus, prior research typically has focused on (last-round) behavior in repeated games. Arguably the most important reason for why a Nash-concept may not be suited to behavior 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 behavior well.

Research Strategy: If the RPNE was meant to describe what actually goes on between participants in a public-good experiment, a number of conditions would have to be fulfilled:

- stability of preferences: whenever $r_i(\mathbf{c}_{-i})$ is elicited, the response should be the same. Thus, both indifference between responses and probabilistic responses would make the RPNE prediction less reliable.
- mutual knowledge of preferences: each player $i \in \{1,...,n\}$ knows all other players' response functions $r_j(\mathbf{c}_{-j}), \forall j \neq i$. If players do not know other players' response functions, it is impossible for them to calculate the RPNE (an RPNE still could be expected to be implemented in case of dominant strategies).

I will refer to these conditions as *pre-conditions for RPNE*. Empirically, in most cases both pre-conditions will be violated, and with respect to those cases, I will be comparing data to an *as-if*-model (mirroring, e.g., the analysis in Section V in Fehr & Schmidt, 1999). In addition, I will be analyzing whether the data correspond to the RPNE prediction more closely when participants' level of preference stability is high, and when mutual knowledge of preferences is induced. If I find positive evidence in either analysis, I will interpret it as indirect support of the RPNE "mechanics," as such findings would suggest that the more the pre-conditions for RPNE apply, the better its predictive power.

TABLE 1 Example for the calculation of an RPNE prediction: The table shows what a row-player of each type would play in the RPNE against the respective column-player type.

	Free-rider	Perfect conditional cooperator	Imperfect conditional cooperator
Free-rider	0 (6.25%)	0 (12.5%)	0 (6.25%)
Perfect conditional cooperator	0 (12.5%)	2 (25%)	1 (12.5%)
Imperfect conditional cooperator	0 (6.25%)	1 (12.5%)	1 (6.25%)

Note: The aggregate prediction would then be that if we subject (new) people to a simultaneous-move public-good experiment using the same parameters, we would observe choices of $\{0, 1, 2\}$ in 43.75%, 31.25%, and 25% of the cases, respectively.

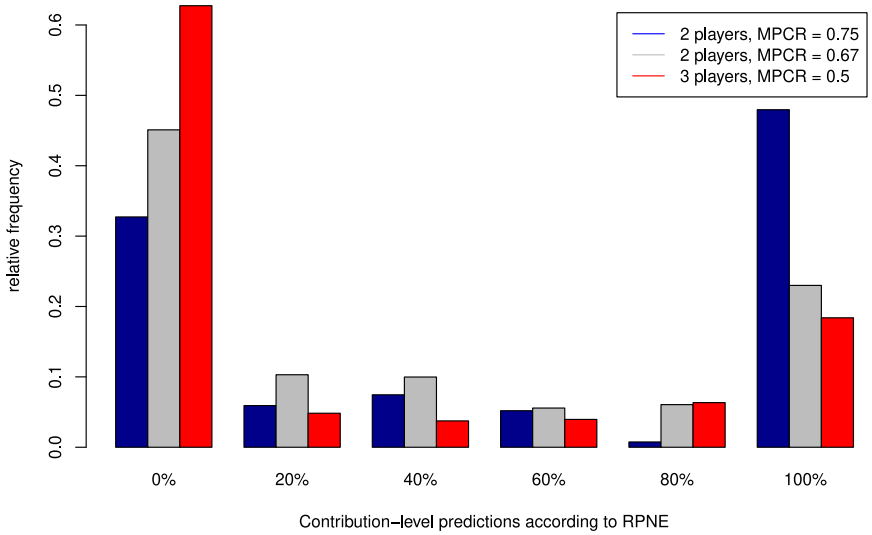

FIGURE 1 RPNE predictions from Wolff (2017).

3 | RESEARCH QUESTIONS

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.[5] Unfortunately, few popular social-preference models come with a calibration that would allow to make such a prediction.[6] 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 behavior (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)?*

A priori, finding a positive answer would be surprising on two accounts. First, the work by Healy (2011) and Brunner et al. (2021) suggested that a Nash-concept based on preference-measurements does not account for public-good behavior. Notably, though, 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 behavior 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.

The second reason for why a positive answer would be 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 play a role?*

To answer *RQ 2*, I conduct an additional "Public Preferences" experiment that creates an environment that approximates mutual knowledge of preferences.[7] The additional experiment provides information on whether the RPNE predicts behavior for the right reasons.

The remainder of the analysis focuses on three goals. First, relating to a discussion in the current prisoner's-dilemma literature, I look at strategic uncertainty. Second, an individual-level analysis of the Public-Preferences data relates the equilibrium pre-conditions to how well participants' behavior can be predicted. And 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?*

4 | MODEL PREDICTIONS

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 calculated and categorized. Specifically, Wolff (2017) calculated the RPNE set for each of his participants when matched with each other participant. Applying the Pareto-dominance criterion then gives a unique contribution profile for each match, and the combination of all possible matches gives the prediction of the relative frequencies with which each contribution level will be played. The calculation is done for a three-player situation with a MPCR μ of μ 0.5, and for two-player situations with μ 2/3 and μ 0.75 (the elicitation of the underlying conditional-contribution preferences follows a procedure that is very similar to the PREFS-experiment I describe in Section 5.1).

In addition to calculating the RPNE sets, 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). In their model, agents have a utility function

$$u_i(\mathbf{x}) \quad x_i - \alpha_i \cdot \frac{1}{n-1}\sum_{j\neq i}\max\{x_j - x_i, 0\} - \beta_i \cdot \frac{1}{n-1}\sum_{j\neq i}\max\{x_i - x_j, 0\},$$

where $\mathbf{x}$ is the vector of monetary payoffs, α_i measures i's dislike of disadvantageous inequality, and β_i measures i's dislike of advantageous inequality. Their calibration of the model (together with their assumption that $\alpha_i \geq \beta_i$) yields four (α_i, β_i) types: 30% (0, 0)-types, 30% (0.5, 0.25)-types, 30% (1, 0.6)-types, and 10% (4, 0.6)-types (see Table III in Fehr & Schmidt, 1999). Using the relative frequencies of the four types then allows to calculate the prediction of the relative frequencies with which each contribution level will be played (taking into account the number of players and μ).[8] As an example, for the 3-player setting with a MPCR of μ 0.5, the model has 94% of agents contributing nothing and 6% contributing everything.

The general upshot is that the RPNE predicts positive contributions substantially more often than Fehr and Schmidt (1999, for example, 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 behavior. Thus, while Wolff (2017) introduced the concept, this paper contributes by examining whether, when, and why the RPNE predicts actual behavior.

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 behavior). However, the learning part does not apply to one-shot settings, which is why I only consider the social-preference part of their model that is meant to account for unexperienced play.

In their model, agents have a utility function

$$u_i(\mathbf{x}) \quad x_i + \delta_i \cdot \frac{\sum_{j \ 1}^{n} x_j}{n} - \gamma_i \cdot \max\left\{ \frac{\sum_{j \ 1}^{n} x_j}{n} - x_i, 0 \right\},$$

where $\mathbf{x}$ again is the vector of monetary payoffs, $\delta_i \geq 0$ measures i's preference for efficiency, and $\gamma_i \geq 0$ measures i's dislike of disadvantageous inequality with respect to the average payoff in the group. They then estimate the distribution of (δ_i, γ_i) types to be 48% (0,0)-types, and the remaining population to follow a distribution $F(\delta_i, \gamma_i) \quad \mathrm{U}([0,2]) \times \mathrm{U}([0,8])$. Most importantly, the model of Arifovic and Ledyard (2012) is able to account for contributions that are neither 0 nor participants' full endowment. As an example, for the 3-player setting with a MPCR of $\mu \quad 0.5$, the model has 55% of agents contributing nothing, 32% contributing 20% of their endowment, and 13% contributing everything.

Finally, research question *RQ 2* parallels current discussions in the literature on indefinitely-repeated prisoners'-dilemma experiments. In particular, Boczoń et al. (2024) analyze (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 behavior 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.

5 | 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., 2010; Drouvelis et al., 2015; Dufwenberg et al., 2011).[9] The three-player studies all used marginal per capita returns of $\mu \quad 0.5$, while the two-player studies had different μs (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: The STANDARD Experiment is a standard one-shot simultaneous linear public-good experiment, with a $\mu \quad 2/3$ and contribution levels of $\{0, 3, 6, ..., 15\}$ "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.[10]

The PUBLIC-PREFERENCES Experiment: The PUBLIC-PREFERENCES Experiment is a more complicated experiment consisting 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.

The focus of the PUBLIC-PREFERENCES Experiment is on the predictability of contribution behavior in the "SIMPG task" and on how this predictability depends on the RPNE pre-conditions detailed at the end of Section 2. The SIMPG task is a standard two-player one-shot linear public-good experiment, except for the fact that participants see their interaction partner's responses from an earlier "PREFS1" part. The PREFS1 part is a standard elicitation of conditional-contribution preferences. I assess the individual-level predictability of participants' contribution behavior in Section 6.2 by contrasting the SIMPG-part choices to the RPNE predictions that result from the PREFS1 measurements.

As outlined at the end of Section 2, 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 (PREFS1, PREFS2, and PREFS3).[11] And to assess pre-condition (ii), a "STABILITYBELIEFS

task" elicits participants' probabilistic beliefs about others' behavior in the PREFS2-part showing them these others' behavior from the PREFS1-part.

In the following, I provide an overview of the seven parts of the experiment in their order of appearance in a session, with a short description of each of them. I follow up with a more detailed description of the focal parts, referring the interested reader to Appendix B in Supporting Information S1 for the exact details of the remaining parts.

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 6.2 (for each randomly-formed pair of participants, one of the 13 dictator decisions as well as the role-assignment are determined randomly at the end of the experiment).

PREFS1. A standard elicitation of conditional-contribution preferences ("PREFS task," Fischbacher et al., 2001), detailed in Section 5.1. My empirical measurement of the response functions $r_i(\mathbf{c}_{-i})$.

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 & Page, 1990; Hossain & Okui, 2013, probability of receiving a prize of 2 Euros determined by a quadratic scoring rule; I do not analyze the beliefs from this part).[12]

SIMPG. The focal simultaneous public-good interaction also detailed in Section 5.1. This is the action a_i to be predicted.

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). Used only for explorative purposes to understand behavior better but not in any of the predictions. Corresponds to the beliefs b_i in Gächter et al.'s (2017) notation.

STABILITYBELIEFS. Elicitation of beliefs in others' elicited-preference stability, with respect to the SIMPG-opponent and three randomly-chosen others: Participants saw another participant's response vector from the PREFS1 part. Then, they had to state a probabilistic belief on the response-vector of the same other participant from the PREFS2 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 PREFS2 (i.e., they had to specify 6×6 probabilities for each of the four others). For each of the four others whose behavioral stability participants had to assess, one first-mover contribution was randomly drawn. Participants were incentivized by a binarised scoring rule for their belief accuracy in the four randomly-drawn cases, with a prize of 6 Euros per lottery.

PREFS3. Final repetition of the PREFS task with a new interaction partner.[13]

Note on signaling incentives in the PUBLIC-PREFERENCES Experiment: Note that if participants know that their behavior in one task may be revealed to others in a later task, they may have potential signaling incentives in the first task.[14] My experimental design addresses the signaling 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 signaling prohibitively costly). They do not get any information about others' behavior 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 signaling attempts by participants, I could not find any evidence of signaling in the data.

5.1 | Specifications of the SIMPG- and the PREFS-tasks

The SIMPG-task consists of a simultaneous two-player linear public-good situation with an MPCR of $\mu \quad \frac{2}{3}$ 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-task, participants face the same two-player linear public-good payoff structure with an MPCR of $\mu \quad \frac{2}{3}$ and an endowment of 15 Euros as in the SIMPG-task. However, the PREFS-experiment differs from the SIMPG in that there is no information on the other player, and in that the PREFS-tasks 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.[15]

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 PREFS-experiments, so that participants would know exactly what payoff profile each of their actions meant. As a third measure, I recruited *experienced* participants.[16] Participants in the experiment had participated in at least one public-good experiment and at least four additional other experiments, with no upper limits.

5.2 | 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. 72 participants from the LakeLab's participant pool of university students took part in the experiment, earning about 13.80 Euros (USD 16.60) on average for about 1 hour (including the unrelated second part mentioned in footnote 10). The earnings include a show-up fee of 5 Euros.

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 cubicle. 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 in Supporting Information S1.

Participants earned on average 19.33 Euros (USD 22) for about 90 min; 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 6.1. For the individual-level analyses in Section 6.2, I then use the data from all seven sessions (PUBLIC PREFERENCES-ALL).

6 | RESULTS

I structure the results section into three parts. In Section 6.1, 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.[17] This part is a test of the different models' external validity and penalizes the calibrated models in case of over-fitting.

Section 6.2 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. Specifically, for the individual-level RPNE prediction for the PUBLIC-PREFERENCES Experiment, I determine who is playing whom in the SIMPG part. Then, I calculate these participants' RPNE sets using (only) their choices from the PREFS1 part and compare a participant's contribution from the Pareto-dominant RPNE to the participant's SIMPG contribution. For the Fehr-Schmidt and Arifovic-Ledyard models, I proceed analogously, only that I use a maximum-likelihood estimate of each participant's model parameters from the participant's choices in the SVO part.

Finally, in Section 6.3, 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.

Note on player-type categories in PUBLIC PREFERENCES: Relating to the RPNE pre-conditions discussed at the end of Section 2, I categorize participants into 2 × 2 categories. In the following paragraphs, I outline the categories and specify the corresponding criteria. I categorize 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.[18] Participants who violate the criterion are categorized 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 categorized as having *incomplete information* with respect to others' preferences or conforming to *mutual knowledge* (of preferences) based on the participant's STABILITYBELIEFS. In the STABILITYBELIEFS part, each participant sees the choices of four other participants from the PREFS1 part and has to state a probablistic belief about the four others' choices in the PREFS2 part.

For the *incomplete-information/mutual-knowledge* categorization, 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.[19] I categorize 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.[20]

The above typology partitions the population into four groups with the following relative frequencies: *consolidated preferences/mutual knowledge*: 30%; *floating preferences/mutual knowledge*: 26%; *consolidated preferences/incomplete information*: 13%; and *floating preferences/incomplete information*: 31%.

6.1 | Out-of-sample predictions

As a measure for the models' predictive power, I use their mean squared prediction errors. To calculate them, I first take the difference between the model's predicted percentage choosing each particular action with the percentage actually observed for that action. The resulting six differences are taken to the power of two, and averaged across all actions.

Table 2 reports the mean squared prediction errors of the out-of-sample RPNE predictions for the eight data sets (where the predictions come from Wolff, 2017).[21] Note that Kamei (2016) and Blanco et al. (2011) use marginal per capita returns (μ 0.6 and μ 0.7, respectively) for which I do not have an RPNE prediction. I use the predictions for μ 2/3 for these two data sets, arguing that the μs are sufficiently close to yield similar results.[22]

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; both as calibrated by the original authors). The prediction of the calibrated Levine (1998) model coincides with "selfish Nash." For reasons of comparability, I adopted the Pareto-dominance criterion from Fehr and Schmidt (1999) also for the RPNE prediction in case of multiple equilibria.

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 versus 0.0120).[23] 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 (μ 0.7 for Blanco et al., 2011, μ 0.6 for Kamei, 2016) from the data the prediction was based on (μ 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).[24] 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 2 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."[25] Second, the RPNE's predictive power is particularly strong for those participants of the STANDARD treatment who report low strategic uncertainty. To measure subjective strategic (un-)

TABLE 2 Mean squared prediction errors of the stated models for the different data sets (the prediction of Levine's (1998), coincides with "selfish Nash"). Smallest mean squared prediction errors are marked in boldface.

Data	"Selfish Nash"	FS (1999)	AL (2012)	Wolff (2017) [a]
Kamei (2016; $n = 2, \mu = 0.6; N = 300$)	0.0761	0.0484	0.0575	**0.0080**[b]
Blanco et al. (2011; $n = 2, \mu = 0.7; N = 72$)	0.1042	0.0653	0.0181	**0.0104**[b]
Guala et al. (2013; $n = 2, \mu = 0.75; N = 410$)	0.1384	0.0615	0.0451	**0.0360**
Cubitt et al. (2010; $n = 3, \mu = 0.5; N = 87$)	0.0613	0.0374	**0.0101**	0.0145
Drouvelis et al. (2015; $n = 3, \mu = 0.5; N = 150$)	0.0736	0.0470	**0.0190**	0.0192
Dufwenberg et al. (2011; $n = 3, \mu = 0.5; N = 303$)	0.0454	0.0247	0.0090	**0.0068**
STANDARD ($n = 2, \mu = 2/3; N = 72$)	0.1191	0.0757	0.0269	**0.0111**
High strategic uncertainty	0.1632	0.1100	0.0468	**0.0255**
Low strategic uncertainty	0.0824	0.0488	0.0144	**0.0043**
PUBLIC PREFERENCES-NEW ($n = 2, \mu = 2/3; N = 70$)	0.0702	0.0384	0.0139	**0.0020**
Floating preferences, incomplete information	0.1136	0.0716	0.0207	**0.0108**
Consolidated preferences, incomplete information	0.0473	0.0313	0.0240	**0.0132**
Floating preferences, mutual knowledge	0.0567	0.0297	0.0168	**0.0035**
Consolidated preferences, mutual knowledge	0.0761	0.0376	0.0152	**0.0022**
Weighted mean squared prediction error over all 8 data sets	0.0893	0.0481	0.0317	0.0171
p-value, Wilcoxon signed-ranks test against Wolff (2017); $N = 8$	0.008	0.008	0.055	–

[a]In case of multiplicity, I adopt Fehr and Schmidt's (1999) Pareto-dominance criterion.

[b]Prediction for $n = 2; \mu = 2/3$.

certainty in the STANDARD treatment, I calculate the sum of squared deviations of the participants' action-beliefs from a uniform distribution. The idea is that this measure is smallest when participants absolutely do not know what their opponent will be doing, and largest when they feel they know it exactly. 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 data row of Table 2, 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 9th data row of Table 2.[26] 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 much 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 (and show that it nevertheless does matter in the expected direction).

Figure 2 shows a histogramme for the RPNE prediction and the data from the STANDARD and PUBLIC-PREFERENCES experiments, to obtain an idea of where the predictions fail. Figure 2 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 no definite conclusions can be drawn 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

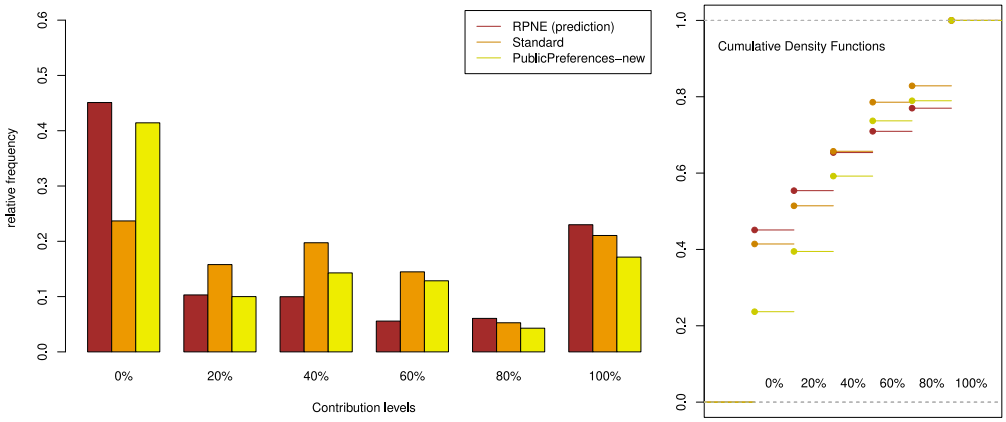

FIGURE 2 Histogramme (left) and cumulative densities (right) for the RPNE prediction from Wolff (2017) and the data from the STANDARD and the PUBLIC PREFERENCES-NEW experiments.

advantage of demonstrating external validity and penalizing over-fitting. On top, they can be tested even when the assumptions of the model are violated 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 behavior reflects the modeled 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.[27]

6.2 | 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 behavior in a simultaneous public-good task using elicited conditional-contribution preferences, this time for individual RPNE predictions. The individual-level analysis differs from the out-of-sample approach particularly in that, in the individual-level analysis, there is a specific predicted contribution level for each participant.

Table 3 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.[28]

To calculate the hit rates in Table 3 for the models by Fehr and Schmidt (1999) and Arifovic and Ledyard (2012), I first estimated participants' individual model parameters from the various dictator games in the SVO-part of the experiment, 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 and Schmidt's (1999) Pareto-dominance criterion and ignored all cases in which the model is mute (because the RPNE set is empty).[29]

Table 3 shows a number of things. First, all of the models clearly are better than a uniform-randomization 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 2, 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 3 shows that the RPNE model does better in predicting individual behavior 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 pre-conditions seem to be fulfilled. For the subset of participants whose preferences seem to be "consolidated" and who generally think that the PREFS1-responses reflect others' preferences, the (Pareto-dominant) RPNE exactly predicts about two thirds of all choices.

TABLE 3 Hit rates for individual-level predictions of the stated models for the PUBLIC PREFERENCES-ALL data (in %).

	"Selfish Nash"	FS (1999)	AL (2012)	RPNE[a]
PUBLIC PREFERENCES-ALL ($N = 152$)	41	41	29	51
Floating preferences, incomplete information	19	23	19	36
Consolidated preferences, incomplete information	40	40	40	45
Floating preferences, mutual knowledge	49	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	–

[a]In case of multiplicity, I adopt Fehr and 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.

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 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."[30]

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 PREFS1-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.[31] 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."

6.3 | 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 4 gives an answer.

As we can see from the first row of Table 4, the Pareto-criterion is clearly the modal criterion for choices that are consistent with an RPNE prediction (on par with non-equilibrium actions only if I pool the latter). Still, 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.

As a side note, the observations above indicate that introducing decision errors is unlikely to improve the predictive power of the Pareto-dominant RPNE: there are substantial relative frequencies on the average-contribution-sum and the

TABLE 4 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 %).

	Pareto	Minimum or average	Other	Non-equilibrium
PUBLIC PREFERENCES-ALL (cases with multiple RPNE, $N = 50$)	36	26	2	36
Both incomplete-information types ($N = 18$)	28	11	11	50
Floating preferences, mutual knowledge ($N = 16$)	38	31	0	31
Consolidated preferences, mutual knowledge ($N = 16$)	44	38	0	19

Note: For consolidated-preferences/incomplete-information, there were only 3 observations (1 "Pareto," 2 non-equilibrium). I thus combine the incomplete-information categories.

minimum-contribution-sum equilibria. Thus, deviations from the Pareto-dominant predictions are non-random, and often go to contribution levels that are "far away" from the Pareto-dominant prediction.

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 RPNE pre-conditions seem to be fulfilled.[32] This suggests two things: first, that unsuccessful predictions are due only partially to participants using a different equilibrium-selection criterion. Reversing the argument, it can be argued that for the consolidated-preferences/mutual-knowledge type equilibrium miscoordination is to blame, which suggests we cannot assume these participants are making errors, which makes a (pure) quantal-response approach seem unpromising.

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 31% 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.

7 | 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 behavior in one-shot public-good experiments. Both prior research (Brunner et al., 2021; Healy, 2011) and plausibility considerations (participants cannot know each others' preferences in a one-shot environment) would have cast serious doubt on this endeavor a priori. Nonetheless, I show that the RPNE predicts behavior 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 behavior 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 behavior 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 behavior 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). Focusing 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 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. This aligns very well with the findings of the seminal paper by van Huyck et al. (1990), who find that 31% of the first-round choices in their minimum-effort game are consistent with the Pareto-dominant equilibrium.

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. The first remark refers to the question of whether there are different types of participants who differ in their propensity to show equilibrium behavior, or whether there is a single type whose propensity to show equilibrium behavior differs between situations. My understanding is that the heterogeneous-types explanation is the most likely one.

This understanding is based on the fact that the categorization 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).[33] Thus, the RPNEs faced by "equilibrium types" generally also do not differ from those faced by other (more) "non-equilibrium types." And hence, a person's propensity to show equilibrium behavior does not seem to depend on the situation but rather on the person's own characteristics.

The second remark concerns why the "mutual-knowledge" assumption seems to be so important. My favored way to understand the finding goes through best-response behavior. Four fifth of choices that are "approximated best-responses" turn out to be in line with the RPNE prediction for any of the four behavioral 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 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' behavior is stable and predictable. Thus, they trust their expectations about their opponent's behavior 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 behavior than situation B, then the Nash-equilibrium will be more predictive of behavior in situation A, *irrespective* of whether behavior 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.

ACKNOWLEDGMENTS

I am grateful to Malte Baader, 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, as well as to Stefano Barbieri and two anonymous reviewers for their valuable questions and suggestions. 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. I computerized 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 & Millo, 2008) and lmtest (Zeileis & 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 & 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 (the "non-emulation," not the liquid) for the programming of the experiment. The article was written using Kile. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in OPENICPSR at https://doi.org/10.3886/E213261V2 (Wolff, 2025).

ENDNOTES

[1] 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).

[2] Note that my notation differs from the notation of Gächter et al. (2017) introduced earlier: their a_i is closer to my r_i, only that it takes a belief as its argument, while r_i takes a realized action profile as the argument.

[3] Prior research supports this assumption (Fischbacher et al., 2012).

[4] Wolff (2017) categorized 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 behavior.

[5] 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."

[6] Arifovic and Ledyard (2012), Fehr and Schmidt (1999), and Levine (1998) are notable exceptions.

[7] I will be explicit below about how I deal with the potential signaling 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 5.1. In essence, I use a similar approach as Brunner et al. (2021).

[8] For an extensive description of how to arrive at the model's prediction, see Fehr and Schmidt (1999, p. 845); note that the procedure to arrive at the RPNE is fully analogous.

[9] 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., 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., 2022) 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.

[10] 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.

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

[12] 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 & Okui, 2013).

[13] In contrast to the first two PREFS tasks, the first-mover in PREFS3 was shown the response-vector of the second-mover from the PREFS1 part before deciding on her (unconditional) contribution. However, the situation of the second-mover was exactly the same as in the PREFS1 and PREFS2 parts. For the purpose of this paper, I therefore regard the PREFS3 part simply as a second repeat-measurement of participants' preferences. I did not analyze the PREFS3 first-mover behavior.

[14] To avoid deceiving participants, the instructions included the sentence that "your behavior from one of the earlier parts will possibly be displayed to other participants in a later part."

[15] The order of the combinations was randomized 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.

[16] I nonetheless asked the usual comprehension questions; participants could only proceed to the experiment after answering all questions correctly.

[17] Note that I do not rely on having any training samples, as I explicitly focus on models that come with a calibration or models that do not need any calibration.

[18] 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 11 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.

[19] The predictive power actually is slightly worse when categorizing 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 behavioral 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.

[20] Changing the threshold to, for example, 70% does not change the results in any meaningful way.

[21] Wolff (2017) computes the RPNE predictions from conditional-contribution vectors elicited from a large number of participants.

[22] 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 ($\mu = 2/3$, average: 37%) and those in Blanco et al. (2011; $\mu = 0.7$, average: 48%).

[23] 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).

[24] 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.

[25] 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 10th data row of Table 2 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.

[26] The contrast replicates, albeit not as pronouncedly, in treatment PUBLIC PREFERENCES-NEW, with mean squared prediction errors of 0.0197 versus 0.0092.

[27] 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.

[28] The main qualitative results continue to hold when I use mean squared prediction errors instead.

[29] Counting these cases as "misses" would yield the following percentages: 48, 34, 45, 46, and 65, in the order given in Table 3. 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" behavior in case of an empty RPNE set.

[30] Pair-wise Wilcoxon–Mann–Whitney tests all yield $p \leq 0.04$ except for the comparison between the "intermediate categories" (consolidated-preferences/incomplete-information versus floating-preference/mutual-knowledge, $p = 0.718$).

[31] 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%.

[32] 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.

[33] Unless a participant is "Unclassifiable"; see Appendix A in Supporting Information S1.

REFERENCES

Ambrus, A. & Pathak, P.A. (2011) Cooperation over finite horizons: a theory and experiments. *Journal of Public Economics*, 95(7–8), 500–512. Available from: https://doi.org/10.1016/j.jpubeco.2010.11.016

Andreoni, J. & Samuelson, L. (2006) Building rational cooperation. *Journal of Economic Theory*, 127(1), 117–154. Available from: https://doi.org/10.1016/j.jet.2004.09.002

Arifovic, J. & Ledyard, J.O. (2012) Individual evolutionary learning, contributions mechanism: other-regarding preferences, and the voluntary contributions mechanism. *Journal of Public Economics*, 96(9–10), 808–823. Available from: https://doi.org/10.1016/j.jpubeco.2012.05.013

Aumann, R. & Brandenburger, A. (1995) Epistemic conditions for Nash equilibrium. *Econometrica*, 5, 1161–1180. Available from: https://doi.org/10.2307/2171725

Blanco, M., Engelmann, D. & Normann, H.T. (2011) A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, 72(2), 321–338. Available from: https://doi.org/10.1016/j.geb.2010.09.008

Bock, O., Baetge, I. & Nicklisch, A. (2014) hroot: Hamburg registration and organization online tool. *European Economic Review*, 71, 117–120. Available from: https://doi.org/10.1016/j.euroecorev.2014.07.003

Boczoń, M., Vespa, E., Weidman, T. & Wilson, A.J. (2024) Testing models of strategic uncertainty: equilibrium selection in repeated games. *Journal of the European Economic Association*, jvae042. Available from: https://doi.org/10.1093/jeea/jvae042

Brunner, C., Kauffeldt, T.F. & Rau, H. (2021) Does mutual knowledge of preferences lead to more Nash equilibrium play? Experimental evidence. *European Economic Review*, 135, 103735. Available from: https://doi.org/10.1016/j.euroecorev.2021.103735

Calhoun, P. (2015) Exact: unconditional exact test. R package version, 1.6.

Croissant, Y. & Millo, G. (2008) Panel data econometrics in R: the plm package. *Journal of Statistical Software*, 27(2). Available from: https://doi.org/10.18637/jss.v027.i02

Cubitt, R.P., Drouvelis, M. & Gächter, S. (2010) Framing and free riding: emotional responses and punishment in social dilemma games. *Experimental Economics*, 14(2), 254–272. Available from: https://doi.org/10.1007/s10683-010-9266-0

Drouvelis, M., Metcalfe, R. & Powdthavee, N. (2015) Can priming cooperation increase public good contributions? *Theory and Decision*, 79(3), 479–492. Available from: https://doi.org/10.1007/s11238-015-9481-4

Duch, M.L., Grossmann, M.R. & 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. Available from: https://doi.org/10.1016/j.jbef.2020.100400

Dufwenberg, M., Gächter, S. & Hennig-Schmidt, H. (2011) The framing of games and the psychology of play. *Games and Economic Behavior*, 73(2), 459–478. Available from: https://doi.org/10.1016/j.geb.2011.02.003

Falk, A. & Fischbacher, U. (2006) A theory of reciprocity. *Games and Economic Behavior*, 54(2), 293–315. Available from: https://doi.org/10.1016/j.geb.2005.03.001

Fehr, E. & Schmidt, K.M. (1999) A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3), 817–868. Available from: https://doi.org/10.1162/003355399556151

Fischbacher, U. (2007) z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2), 171–178. Available from: https://doi.org/10.1007/s10683-006-9159-4

Fischbacher, U. & Gächter, S. (2010) Social preferences, beliefs and the dynamics of free riding in public good experiments. *The American Economic Review*, 100(1), 541–556. Available from: https://doi.org/10.1257/aer.100.1.541

Fischbacher, U., Gächter, S. & Fehr, E. (2001) Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3), 397–404. Available from: https://doi.org/10.1016/s0165-1765(01)00394-9

Fischbacher, U., Gächter, S. & Quercia, S. (2012) The behavioral validity of the strategy method in public good experiments. *Journal of Economic Psychology*, 33(4), 897–913. Available from: https://doi.org/10.1016/j.joep.2012.04.002

Gächter, S., Kölle, F. & Quercia, S. (2017) Reciprocity and the tragedies of maintaining and providing the commons. *Nature Human Behaviour*, 1(9), 650–656. Available from: https://doi.org/10.1038/s41562-017-0191-5

Greiner, B. (2004) An online recruitment system for economic experiments. In Kremer, K. & Macho, V. (Eds.) *Forschung und wissenschaftliches Rechnen 2003*. Göttingen: Ges. für Wiss. Datenverarbeitung, vol. 63, pp. 79–93.

Guala, F., Mittone, L. & Ploner, M. (2013) Group membership, team preferences, and expectations. *Journal of Economic Behavior & Organization*, 86, 183–190. Available from: https://doi.org/10.1016/j.jebo.2012.12.003

Healy, P.J. (2011) Epistemic foundations for the failure of Nash equilibrium. Working paper.

Højsgaard, S. & Halekoh, U. (2016) doBy: Groupwise Statistics, LSmeans, Linear Estimates, Utilities. *R package version*, 4, 5–15.

Hossain, T. & Okui, R. (2013) The binarized scoring rule. *The Review of Economic Studies*, 80(3), 984–1001. Available from: https://doi.org/10.1093/restud/rdt006

Ihaka, R. (1998) R: past and future history. In Weisberg, S. (Ed.) *Proceedings of the 30th Symposium on the Interface*, pp. 392–396.

Kamei, K. (2016) Democracy and resilient pro-social behavioral change: an experimental study. *Social Choice and Welfare*, 47(2), 359–378. Available from: https://doi.org/10.1007/s00355-016-0967-y

Kartal, M. & 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. Available at: http://www.kde.org/community/whatiskde/ [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. Available from: https://doi.org/10.18637/jss.v055.i08

Levine, D.K. (1998) Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, 1(3), 593–622. Available from: https://doi.org/10.1006/redy.1998.0023

McKelvey, R.D. & Page, T. (1990) Public and private information: an experimental study of information pooling. *Econometrica*, 58(6), 1321–1339. Available from: https://doi.org/10.2307/2938318

Murphy, R.O., Ackermann, K.A. & Handgraaf, M.J.J. (2011) Measuring social value orientation. *Judgment and Decision Making*, 6(8), 771–781. Available from: https://doi.org/10.1017/s1930297500004204

Offerman, T., Sonnemans, J. & Schram, A. (1996) Value orientations, expectations and voluntary contributions in public goods. *The Economic Journal*, 106(437), 817–845. Available from: https://doi.org/10.2307/2235360

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.

Rapoport, A. (1985) Provision of public goods and the MCS experimental paradigm. *American Political Science Review*, 79(1), 148–155. Available from: https://doi.org/10.2307/1956124

Rapoport, A. & Suleiman, R. (1992) Equilibrium solutions for resource dilemmas. *Group Decision and Negotiation*, 1(3), 269–294. Available from: https://doi.org/10.1007/bf00126266

Rödiger, S., Friedrichsmeier, T., Kapat, P. & 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. Available from: https://doi.org/10.18637/jss.v049.i09

RStudio Team. (2015) *RStudio: Integrated Development Environment for R*. Boston: RStudio, Inc.

Spadaro, G., Tiddi, I., Columbus, S., Jin, S., Ten Teije, A., CoDa Team, T., et al. (2022) The cooperation databank: machine-readable science accelerates research synthesis. *Perspectives on Psychological Science*, 17(5), 1472–1489. Available from: https://doi.org/10.1177/17456916211053319

van Huyck, J.B., Battalio, R.C. & Beil, R.O. (1990) Tacit coordination games, strategic uncertainty, and coordination failure. *The American Economic Review*, 80(1), 234–248.

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. Available from: https://doi.org/10.1016/j.econlet.2016.11.015

Wolff, I. (2025) *ECIN replication package for "Predicting voluntary contributions by 'revealed-preference Nash equilibrium.'"*. Ann Arbor: Interuniversity Consortium for Political and Social Research [distributor]. Available from: https://doi.org/10.3886/E213261V2

Zeileis, A. & Hothorn, T. (2002) Diagnostic checking in regression relationships. *R News*, 2(3), 7–10.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wolff, I. (2025) Predicting voluntary contributions by "revealed-preference Nash-equilibrium". *Economic Inquiry*, 63(3), 846–864. Available from: https://doi.org/10.1111/ecin.13280