Validation of the Conditional Cooperation Cognitive Model for Public Goods Games

Course: Decision Making, E23

Professor: Andreas Højlund Lorenzen

January 10th, 2024

MSc Cognitive Science
Aarhus University
School of Communication and Culture

Authors:

[redacted]

AARHUS UNIVERSITY

Table of contents

Abstract [redacted] 4

1. Introductio [redacted] 5

1.1 The Public Goods Game... 5

1.1.2 Modelling the Public Goods Game...7

1.1.3 Contribution Tables & Effects of Conditional Cooperation...................................7

1.1.3.1 Classification of Players in Grandjean et al. (2022)....................................... 8

1.2 Cognitive Modelling of Public Goods Games.. 8

1.2.1 Conditional Cooperation in a Cognitive Modelling Framework.............................. 9

1.2.2 Validation of the CC model...10

1.2.3 Exploratory Group Comparisons.. 11

1.3 Hypothesis Formation... 11

1.3.1 CC Model Validation... 11

1.3.2 Group Analysis.. 12

2. Methods...13

2.1 The Dataset [redacted] 13

2.2 Modelling Specifications [redacted] 13

2.2.1 Parameters of Interest...14

2.2.2 Priors...14

2.2.3 Parameter Estimation..15

2.2.4 Deriving β from Subjects' CTs.. 16

2.2.5 Spearman's Rank Correlation Test..16

2.2.6 Parameter Recover [redacted] 17

2.3 Group Comparison: Strategic Ability [redacted] 17

2.3.1 Hierarchical CC Model for Group Analysis.. 18

2.3.2 Parameter Recovery..18

2.3.3 Parameter Estimation & Group Comparison...19

3. Results...19

3.1 CC Model Validation.. 19

3.1.1 Parameter Recovery [redacted] 19

3.1.2 Parameter Estimatio [redacted] 20

3.1.2.1 Convergence Diagnostics..20

3.1.3 Correlation Analysis [redacted] 22

3.2 Group Analysis [redacted] 23

3.2.1 Parameter recovery..23

3.2.2 Parameter estimation...24

Convergence Diagnostics... 24

4. Discussion..24

4.1 Interpretation of Results...25

4.1.1 H1: CC Model Validatio [redacted] 25

4.1.2 H2 & H3 : Exploratory Group Comparison [redacted] 26

4.2 Implications of Results..27

4.3 Limitations and Future Research..29

4.3.1 Deriving β Using Linear Regression [redacted] 29

[redacted]

4.3.2 Model Fitting For Validatio [redacted] 30
4.3.3 Limitations of Cognitive Mo e ng [redacted] 30
4.3.4 Conditional Cooperation Models and Ecological Validit [redacted] 31
5. Conclusion [redacted] **31**
6. Bibliograp y **33**

Abstract[redacted]

In public goods games (PGG), filling out contribution tables (CTs) prior to the game have traditionally been used to derive the cooperation preferences of players, grouping them into categories such as freeriders or conditional cooperators. The present paper aims to validate the cognitive model, and in particular the model parameter ρ (readiness to cooperate), presented by Skewes & Nockur (2023) as a formalisation of the conditional cooperation schema by Fischbacher & Gächter, (2010). It is herein claimed that the cognitive model named the CC model would pass validation if inferred ϱs from in-game behaviour could replace the practice of using the separate CTs to identify preference types. A positive correlation was found between inferred ρs from the CC model and βs from CTs (Rho_{corr} = 0.35, p < 0.0001*), indicating that while there is an association, it is for this specific dataset not high, indicating discrepancy between explicit/implicit preferences for cooperation. The CC model on the individual level achieved satisfactory parameter recovery and convergence. Additionally we investigate the utility of a hierarchical CC model in a group comparison. The model indicated little to no support of differences in attention to others (ω) and ρ for high- vs low-strategy matched players in groups. However, since the dataset is small and hints of directions were seen (indicating high-strategy matched players to, on group level, have slightly higher means of both ω and ρ), this analysis should be replicated with a model achieving better parameter recovery. Implications of the results and limitations are discussed, as well as cases for future research.

Keywords: Conditional cooperation, public goods games, cognitive modelling

All code used in the present study can be found on Github via [redacted]

[redacted]

1. Introduction [redacted]

Human decision making is a complex process. To arrive at a decision, many cognitive processes must interact to sort through possible courses of action - and choose one. Traditional economic theory predicts this 'one' course of action by the desire to maximise utility for one-self, assuming that individuals are rational decision makers seeking personal gain. However, evidence from experiments in behavioural economics with social dilemmas such as the public goods game (PGG) indicates that real human decision making often deviates from these predictions.

Individuals' decisions in PGGs also reflect varied social preferences, contributing more to the collective good than traditional theories suggest, for instance showing the social preferences to condition contributions on what others give or what they *believe* others give. This effect is called *conditional cooperation,* proposed by Fischbacher et al. (2001). The present study sets out to validate Skewes & Nockur's (2023) formal implementation of the conditional cooperation schema presented in Fischbacher et al. (2001, 2010).

The model, the Conditional Cooperation model (henceforth CC model), implements both aspects of social belief and social learning processes to model preferences for conditional cooperation and facilitates extraction of a parameter for readiness to cooperate (ρ) which is for each player based on a cognitive model.

As such, the present study examines the research question of whether the CC model and inferred ρ parameters provide a good alternative to using contribution tables to infer preferences for cooperation in the PGG. The model is also used for a group comparison.

1.1 The Public Goods Game

Economic games facilitate studying decision making and cooperation. One branch of these are concerned with behaviour occurring in social dilemmas, e.g., the public goods game (PGG) (Kagel & Roth, 1995). A player in the PGG starts with an amount of tokens of monetary value. In each round of the game, all players simultaneously and independently decide what to contribute of their tokens to a common pot. This contribution can range from nothing at all to all of their tokens (i.e. their endowment). After each round, the contents of the common pot is multiplied by a constant (>1<n, n being number of players (Thielmann et al., 2021)) and then evenly distributed among players in a group. Herein lies the social dilemma: players who contribute nothing or small amounts can potentially benefit a lot *if*

[redacted]

others contribute more. At the same time, benefits are maximised if everyone contributes maximally. As such, the financial incentive to contribute as few tokens as possible competes with the potential of getting more if everyone “agrees” to do so.

The PGG can be finitely or infinitely repeated, respectively being scenarios where people know/do not know how many rounds are left, which can affect contributions (Lugovskyy et al., 2017). The PGG also sometimes includes a *punishment-condition*, which has shown to reveal other aspects of behaviour than the classic PGG is able to (Dong et al., 2016).

The purpose of games targeting social dilemmas like the PGG is to, in controlled conditions, illuminate the tension between individual self-interest and the collective good. As such, the PGG aims to bring certain decision behaviours to light under circumstances that aim to model behaviours occurring in real life situations (Thielmann et al., 2021). The PGG affords a decision-making arena with an interplay of both individualistic motives, social norms, movements of reciprocity, inequity aversion, altruism, and cooperation (Houser & McCabe, 2014). Understanding social preferences in these situations can elucidate *why people do what they do* and illuminate deviations from standard theories.

Generalised to PGGs, traditional economic theory predicts that players being mostly self-interested and rational in an economic sense would prefer to free-ride on others’ contributions (explaining the problem of freeriders (Ozono et al., 2016)). However, PGGs often yield empirical evidence that shows a much higher degree of cooperation and decision making that reflect other motivations than personal gain. For instance, research suggests that people’s decisions in the PGG can be motivated by a preference for *fairness* (Jeung et al., 2016). Notably, self-interest in economics is defined as driving actions that elicit personal benefit, such as gaining money (Egashira et al., 2021).

Proposed types of players include freeriders, *perfect altruists,* (or unconditional cooperators) who, while rare (Thöni & Volk, 2018), contribute all of their tokens all of the time, and *conditional cooperators,* proposed first by Fischbacher et al., (2001). These are players that condition their contributions based on what others do; if others contribute a lot, so do they, and vice versa for lower contributions. Player (or preference) types are traditionally identified as per the social preferences exhibited in their contribution tables, usually filled out prior to the actual game, as per the ‘strategy method’ (Fischbacher et al., 2001; Selten, 1967), see section 1.1.3.

1.1.2 Modelling the Public Goods Game

A wide palette of cognitive and behavioural models for decision making specifically for the PGG have been presented to explain the processes the PGG seems to provoke or uncover. Skewes & Nockur (2023) review two main approaches. These are a) based on reinforcement learning, where players adjust preferences for contribution depending on feedback (Camerer & Hua Ho, 1999) and b) belief- and social learning based, where players adjust contribution preferences based on *belief* about others' contributions (Fischbacher & Gächter, 2010; Larrouy & Lecouteux, 2017). In Skewes & Nockur (2023), a formalisation of the conditional cooperation (CC) schema from Fischbacher & Gächter (2010) is used to investigate whether inequality (in players' home countries) affects PGG decisions. The current paper addresses the second school of thought in Skewes & Nockur (2023), the belief- and social learning based.

The PGG has functioned as a guide for establishing theories for social preferences, i.e. the degree to which individual choices in economic situations can be influenced by the welfare of others (Houser & McCabe, 2014). Parameters in the PGG are adjustable; variations being possible along e.g. for groupsize (Pereda et al., 2019), group composition devised by matching players that are alike (Grandjean et al., 2022), gender composition in groups (van Staveren et al., 2015), or having changing groups with dropouts and newcomers (Otten et al., 2022).

1.1.3 Contribution Tables & Effects of Conditional Cooperation

The literature shows a consolidated practice for using contribution tables (CTs), to inform a classification of individuals into certain preference types varying in their degree of cooperativeness, by extracting social preferences from the CTs (Houser & McCabe, 2014; Kurzban & Houser, 2005; Thöni & Volk, 2018). The criteria used to classify individuals can vary, but frequencies of player types have been found (gauged via a refined scheme by Thöni & Volk (2018) to overall be stable relative to those presented in Fischbacher et al. (2001).

Finding that cooperation in PGGs tended to decline over time, Fishbacher et al. (2001) first had subjects in a PGG complete a one-shot PGG (the 'unconditional contribution'). Second, subjects were to fill out the aforementioned CTs to measure subjects' preference for conditional cooperation. For the filling out of CTs, each subject was given an empty table with a list of 21 possible average contribution levels of other group members (from 0 to 20 tokens). Subjects had to then indicate for each level how much they would contribute to the common pot, and the filled out table was then used to infer social preferences.

The authors conclude that the reason for the decline in cooperation over time in PGGs is that i) a substantial amount of the players are freeriders and ii) even though the sample here has players that can be classified as conditional cooperators, the majority of them have a bias to often act more according to immediate self-interest, leading to undermatching instead of matching of contributions. The authors also discuss that the speed of convergence of the decline throughout a PGG may depend on group composition, adding "*Positive and stable contributions to the public good are very unlikely*" (Fischbacher et al., 2001, p. 9). A later study by Fischbacher & Gächter (2010) has hinted at the fact that CTs filled out post-gameplay may be more accurate to actual game behaviour compared to pre-gameplay. These potential sequence effects are discussed herein.

1.1.3.1 Classification of Players in Grandjean et al. (2022)

The data used herein from Grandjean et al. (2022) classifies subjects according to their CTs as follows:

- **Freeriders**: If average entry is below 10% of endowment
- **Unconditional Cooperator**: If average entry is higher than average contributions and standard deviation of contributions is below 5% of endowment
- **Conditional Cooperator**: If correlation between entries and corresponding average contribution of others is >0.7 and subject is not a Freerider or an Unconditional Cooperator
- **Others**: If none of the above

Grandjean et al. (2022) study two main explanations proposed for the decline in cooperation in PGGs (Chaudhuri, 2011; Kagel & Roth, 1995): First, the preference-based, which explains the empirical patterns as being due to the interaction between freeriders (who do not contribute) and conditional cooperators (who match others' contributions). Second, the strategy-based, which posits that subjects have stronger incentives to contribute more in the beginning of the game because it may fuel later cooperation.

1.2 Cognitive Modelling of Public Goods Games

Results from PGG have typically been considered in light of traditional statistical tests predicting contributions from observable variables such as economic status, group size, group compositions, etc. While this can provide insight into social behaviour under different conditions, it is not able to take into account fully the underlying drivers – the true 'why's – of the behaviour.

Exactly this is the aim of *cognitive modelling*, an approach which has revolutionised the field of decision-making by combining well-established economic games with theoretically founded computational models of cognition (Katahira, 2016; Wilson & Collins, 2019).

Cognitive models in decision-making are essentially formalisations of cognitive processes into a set of latent variables and their mathematical relationships with observable variables and each other. The relationship between variables is usually based in established theories of decision-making (Prezenski et al., 2017; Wilson & Collins, 2019) and helps to disentangle the influence of different components of a (set of) cognitive process(es). Inherent to cognitive models is the assumption that each individual will vary in certain parameters or latent 'traits' that influence their observable behaviour which can be inferred via a model fit to data.

1.2.1 Conditional Cooperation in a Cognitive Modelling Framework

While research on PGG presents different schools with different approaches appropriate for cognitive modelling, this paper focuses exclusively on the belief-based social learning approach to contributions. Based on the conditional cooperation framework presented by Fischbacher et al. (2001; Fischbacher & Gächter, 2010), Skewes and Nockur (2023) present the CC model, a formalised cognitive model within a Bayesian framework. The model aims to explain behaviour in PGGs including emergent phenomena such as decline in cooperation over time (Muller et al., 2008), from three inherent traits: i) preference for conditional cooperation, ii) amount of attention paid to others, and ii) the initial belief about others' contributions. Skewes & Nockur, (2023), formalise the model as follows. First, the contribution of individuals s in group g on trial t is modelled as a draw from a Poisson distribution:

(1) $$c_{g,s,t} \sim Poisson(P_{g,s,t})$$

The Poisson is appropriate because the distribution of tokens is a discrete outcome variable whose variance increases with size of contribution (Skewes & Nockur, 2023). The parameter P denotes an individual's latent contribution preference on trial t and is modelled as a function of individual's belief of other group members' contribution $Gb_{g,s,t}$ on the trial:

(2) $$P_{g,s,t} = \rho_{g,s} \cdot Gb_{g,s,t}$$

[redacted]

$\rho_{g,s}$ is the preference parameter. This denotes an individual’s preference for being a conditional cooperator, i.e. readiness to cooperate by matching the contributions of others. ρ is constrained between 0 and 1, with 0 being a perfect freerider, 1 being a perfect conditional cooperator and values in between being degrees of undermatching (Skewes & Nockur, 2023). As no intercept for ρ is modelled, the rare cases where a person is an unconditional cooperator, i.e. contributing a consistently high amount regardless of others’ contributions, are disregarded.

A person’s belief about others’ contributions on a given trial $Gb_{g,s,t}$ follows the learning rule:

$$(3) \qquad Gb_{g,s,t} = (1 - \omega_{g,s}) \cdot Gb_{g,s,t-1} + \omega_{g,s} \cdot Ga_{g,s,t-1}$$

Where $Gb_{g,s,t-1}$ is the belief on previous trial, $Ga_{g,s,t-1}$ is the observed mean contribution of others on previous trial, and $\omega_{g,s}$ denotes degree of weight or attention an individual pays to observed contributions of others. If $\omega_{g,s}$ is high, a person is more sensitive to others’ behaviour and will thus quicker update their beliefs about what other individuals in the game will contribute. On the first trial, the group belief is described:

$$(4) \qquad Gb_{g,s,1} \sim Poisson(\alpha_{g,s})$$

Where $\alpha_{g,s}$ is the inherent belief or optimism about the average contribution of others.

1.2.2 Validation of the CC model

Though the CC model is a formalisation of existing theory, the translation into a computational model necessitates choices that may alter the properties of important concepts. As such, it calls for validation (Hiatt et al., 2022).

One of the fundamental concepts in PGG is the *preference* for conditional cooperation, or matching, which has, as reviewed above, traditionally been inferred from CTs filled out prior posterior to playing the PGG game. In the CC model, the preference for conditional cooperation is incorporated as the latent parameter ρ, which is inferred directly from data (and priors) rather than from a separate task.

[redacted]

The two approaches of inference of preference are fundamentally different. The former relies on an individual's explicit knowledge about their own preferences, and the latter infers preferences implicitly through in-game behaviour. A question thus arises of whether the two are comparable, i.e. whether ρ holds the same or at least comparable information about a person's preference towards conditional cooperation as can be found using a CT.

1.2.3 Exploratory Group Comparisons

If one accepts its premise and boundaries, the CC model provides an opportunity to raise a new set of hypotheses about the relationship between classic underlying traits associated with behaviour in PGGs and other 'personality traits'.

In their paper, Grandjean et al. (2022) investigated the correlation between PGG contributions and subjects' strategic ability as determined from a battery of tasks assessing abilities in e.g. reasoning and planning. The authors found that individuals contributed more in groups consisting of high-strategy individuals compared to low-strategy, indicating a fundamental difference between the two groups. This may be explained by differences in traits addressed by the CC-model.

Grandjean et al. base their hypotheses about influence of strategic ability on PGG decisions on the notion that highly strategic players will be adept at understanding the benefits of cooperation than players with lower strategic ability.

Within the CC model framework, this would imply that high-strategy players will have a higher preference for conditional cooperation compared to low-strategy players. Furthermore, the authors suggest that high-strategy players will respond more to differences in group composition and adapt their behaviour accordingly, suggesting a higher attention to the behaviour of others compared to low-strategy players. These suggestions present an interesting case for an exploratory investigation using the CC model, building on the validation of the CC model and extending its utility to allow for comparisons between groups.

1.3 Hypothesis Formation

1.3.1 CC Model Validation

Using an open source dataset from Grandjean et al., (2022), the present study sets out to validate the formalisation of conditional cooperation as presented by Skewes & Nockur

[redacted]

(2023). We focus specifically on the validation of ρ as an expression of individual readiness to cooperate by correlating

a) inferred ρ for each subject from a fitted JAGS model
b) β for each subject from a simple linear regression on preferred contributions from average contribution of others, as derived from the contribution tables

Even though ρ is the main focus in the validation, other parameters α and ω are also extracted for the purpose of inspecting parameter recovery.

By correlating ρs from the CC model and βs from CTs, which according to the authors should at least be strongly correlated, we test the following hypothesis:

H1: Inferred ρs from a CC model fitted using JAGS will correlate significantly with βs for linear regression models fitted to CTs, predicting preferred contribution from average contribution of others.

1.3.2 Group Analysis

Following the model validation, we conduct an exploratory analysis to investigate the utility of the CC model in conditions existent in the dataset. The analysis focuses on a group comparison between two 'types' of people, namely individuals with high strategic abilities and people with low strategic abilities as categorised by Grandjean et al. (2022).

To do so, the model is adapted to infer a group distribution rather than individual level parameters, assuming that there will be a difference in the underlying distributions of the two types. As per Grandjean et al. (2022), we expect highly strategic individuals to have a higher preference for conditional cooperation, leading to the following hypothesis:

H2: The mean of the underlying group distribution for ρ of $\mu\rho$ will be higher for high- compared to low-strategy individuals.

We also expect highly strategic individuals to be more attentive to behaviour of others to quickly adjust their behaviour to be most beneficial, resulting in the hypothesis:

H3: The mean of the underlying group distribution for ω of $\mu\omega$ will be higher for high- compared to low-strategy individuals.

[redacted]

2. Methods

2.1 The Dataset [redacted]

Data from Grandjean et al., (2022) were used. The main PGG dataset consists of 192 individuals across 64 groups of three, who play the PGG for 15 trials. The data also includes CTs filled out prior to gameplay along with classified player types made by the authors based on the aforementioned rules. Finally, the data includes data on individual players and groups strategy level.

Aiming to examine the effect of group composition, Grandjean et al. (2022) matched players in groups across three conditions: *random* (players are matched randomly), *preference* (e.g., freeriders play with freeriders, conditional cooperators with conditional cooperators, and so on), and *strategy* (high-strategy players are matched with other high-strategy players, and vice versa for low-strategy players). Practically, matching in the latter condition was done by sorting levels of strategy and matching top down. No players appeared in more than one condition nor in more groups, allowing us to fit models on the entire dataset while respecting assumptions of independence. In addition, players were unaware of the matching protocol, meaning beliefs about group composition are not expected to influence results.

In all games, each player was given an endowment of 20 tokens (but 200 tokens in the CTs). The earnings of individual i in a group consisting of individuals i, j, and k given the contributions to the common pot of c_i, c_j, and c_k , is

$$\pi_i = 20 - c_i + 0.6(c_i + c_j + c_k)$$

2.2 Modelling Specifications [redacted]

The overall CC model is constructed in line with Skewes and Nockur (2023) (See *Figure 1* for plate notation) and is analysed using Just Another Gibbs Sampler (JAGS) (*JAGS - Just Another Gibbs Sampler*, 2023). JAGS is a program and language using Markov Chain Monte Carlo (MCMC) simulations for approximating posterior distributions of parameters by sampling and being a domain-specific language used to define the model structure, including priors, likelihood, data, etc.

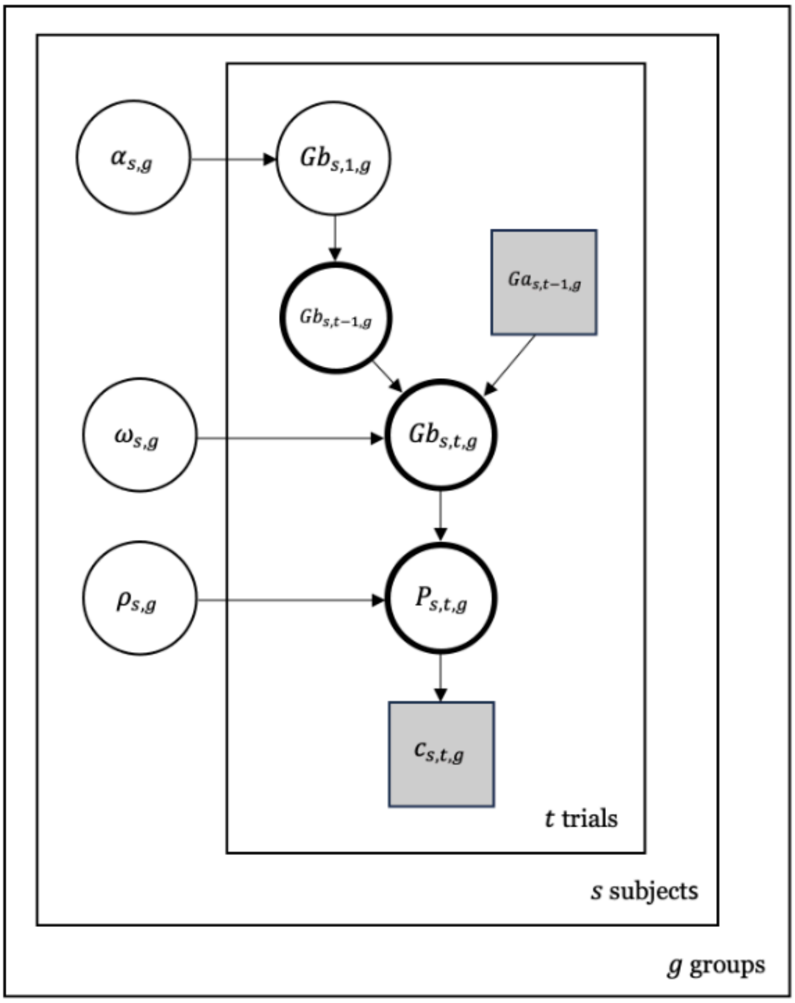


Figure 1. *Plate notation for the CC model. In the present study, the three 'trait' parameters, alpha, omega and rho retrieved.*

2.2.1 Parameters of Interest

To map underlying cognitive processes unfolding in individuals in a PGG, the parameters alpha (α), omega (ω), and rho (ϱ) are extracted. These respectively target modelling α) initial belief about contributions of others, ω) degree of attention to others and ρ) readiness (or preference) to cooperate. Posteriors for these parameters will be inspected and the parameter ϱ is used in the correlation analysis for H1.

2.2.2 Priors

Relatively uninformative priors are set to not nudge results in any specific direction while disallowing impossible values and facilitating convergence. For α, a gamma distribution was

[redacted]

set for the rate parameter used in the subsequent Poisson distribution (a conjugate prior for the Poisson likelihood (McElreath, 2016)). The mean of the gamma distribution is set to 10 (2/0.2 = 10) to constrain possible rates of contribution, setting the average rate of contributions to be around 10 tokens (half of the total). The corresponding variance with these parameters was 50 $(2/0.2)^2 = 50$. The mode of the distribution will be 5 (the actual average contribution in the data was m = 7.25). In sum, we set initial belief about others' contributions to be likely around 5-10 with a fair amount of variability. Provided that the dataset (of substantial size, n= 2880) exhibits a strong signal, the posterior will be pulled towards true values.

For ω and ρ, limits were imposed via truncation on the range of values to respectively remove the extreme values close to 0 and 1 for the Beta(1,1) to avoid values in problematic boundary areas.

In sum, priors were set as follows, the subscript interval denoting truncation limits:

$$\alpha_{g,s} \sim Gamma(2, 0.2)$$

$$\omega_{g,s} \sim Beta_{[0.001,0.999]}(1, 1)$$

$$\rho_{g,s} \sim Beta_{[0.001,0.999]}(1, 1)$$

2.2.3 Parameter Estimation

The CC model specified in section 1.2.1 was fitted in R (R Core Team, 2021) using the R2jags library (Su & Yajima, 2021) and the jags() function, see full code in Appendix. The datalist supplied to jags() consisted of number of groups (64), number of trials (15), the groupsize (3), the average contribution in the group without individual s (a matrix of dimensions 3x15x64), and finally, the actual contribution of player s at trial t for group g (a matrix of dimensions 3x15x64). Matrices were made from the data from Grandjean et al. (2022), using all of the data from all three conditions for validation.

The CC model was specified according to equations in section 1.2.1 to estimate Group Belief, Contribution Preference (from which ρ was extracted as a slope), along with the Contributions as a sample from the estimated Contribution Preference. The model was fitted with 3 chains, 40000 iterations of which 8000 (20%) were burnin, with a thinning interval of 1.

[redacted]

The sampling process was optimised iteratively by inspecting convergence via the Gelman-Rubin (GR) statistic, by visually inspecting trace plots, autocorrelation with lag of 1, and effective sampling (McElreath, 2016).

2.2.4 Deriving β from Subjects' CTs

Similar to in Fischbacher & Gächter (2010), 192 separate simple linear regression models were fitted to each subject's CT using the lm() function in R (R Core Team, 2021), specified as follows for each subject, for observation i:

$$Preferred\ Contribution_i \sim \beta \times Average\ Contribution\ of\ Others_i + \varepsilon_i$$ [1]

This model specification assumes a linear relationship between the preferred contribution and average contribution of other group members with no intercept to match procedure in Fischbacher & Gächter, (2010) and the formalisation in Skewes & Nockur (2023). Post-hoc, models were also fit specifying intercepts (see code) and results of the correlations were robust.

Inspections of the 192 adjusted R^2 values revealed quite a bit of variance in fit, some models scoring 0 (never contributing results in a slope and adjusted R^2 of 0), others scoring nearly a perfectly (i.e. over-) fit. Some CTs also indicated a non-linear spread of contributions, see plots for all models in Appendix. Presence of homoscedasticity of variance was similarly varied across models, see Discussion.

2.2.5 Spearman's Rank Correlation Test

A correlation analysis was run to relate actual contributions in the PGG played in Grandjean et al. (2022) to explicitly stated preferences in CTs. If the correlation was high, this would - under the assumption that CTs provide a good measure of preferences, see Discussion - indicate the cognitive model implemented herein is able to successfully grasp preferences; i.e., filling out CTs as per the 'strategy method' would become redundant as preferences for conditional cooperation can be inferred directly from actual in-game PGG behaviour.

Spearman's Rank Correlation (non-parametric) test (Field et al., 2012) was conducted between the 192 derived βs (i.e., slopes from CTs, normalised between 0 and 1 for interpretability) and the 192 inferred ρs from the cognitive model (in a sense, slopes from the equation for P for subjects' in-game-derived readiness to cooperate), which are also between 0 and 1.

[1] Syntax in R: lm(Preferred_contribution ~ Avg_group_contribution_others + 0, data = df)

[redacted]

For normalised βs, 10 ranks with ties were found. Since the sample size of 192 is relatively large, this number of ranks of 10 (~5% of the data) is modest and deemed unlikely to have a large impact on results. Post-hoc, the correlation was also run i) using Kendall's Tau (which is sometimes better suited for data with ranks) (Field et al., 2012) and ii) non-normalised βs (see code) to assess robustness of results compared to Spearman's with normalised βs. Results were largely the same, see Appendix.

2.2.6 Parameter Recovery [redacted]

As a step in the validation process, we investigate how well the model recovers latent parameters ρ, α, and ω. A simulation was run for 960 groups of three participants over 15 trials, with the CC model directly incorporated as the decision-making process determining the contributions of simulated individuals. For each simulated individual i in each group g, a ρ, α, and ω was sampled as follows:

$$\alpha_{s,g} \sim Uniform(0, 20)$$

$$\rho_{s,g} \sim Uniform(0.001, 0.999)$$

$$\omega_{s,g} \sim Uniform(0.001, 0.999)$$

α was drawn from a distribution between 0 and 20, as the initial belief of others' contribution is constrained within the min/max amounts of tokens that can be contributed. The CC-model specifies that ρ is between 0 and 1 and ω, as a weighting parameter, has the same constraint.

To assess recovery, modes of the posteriors of inferred parameters from a model fitted to the simulated data were plotted against 'true' sampled parameter values for each individual. Additionally, inferred parameters were plotted against other inferred parameters and against other true parameters as a post-hoc assessment of dependencies.

2.3 Group Comparison: Strategic Ability [redacted]

Traits of individuals with high- and low strategic ability in the 'STRAT' condition as presented in the introduction were compared for H2 and H3. All groups from the 'STRAT' condition, including only homogenous groups of either high- or low-strategic individuals as per the classification done in Grandjean et al. (2022), were used: in total, these data comprised 13 'strategy high' groups and 11 'strategy low' groups of low-strategy individuals.

The 'STRAT' condition is chosen to keep the model as simple as possible and to avoid potential confounds as a result of heterogeneous group compositions: by only including this condition we know that high-strategic individuals played the game with other players of similar levels, and vice versa for low-strategic individuals.

2.3.1 Hierarchical CC Model for Group Analysis

The CC model was modified to include a hierarchical element so that each individual's traits are sampled from an underlying distribution with mean μ_{G_i} and standard deviation σ_{G_i}, where G_i denotes category for group of type i, one type being people with high strategic ability (G_{high}) and the other being people with low strategic ability (G_{low}). Both categories have the same specification in the model.

As for the individual model, relatively uninformative priors were set for the group means with similar truncations:

$$\mu^{\alpha}_{G_i} \sim Gamma(0.1, 0.1)$$

$$\mu^{\omega}_{G_i} \sim Beta_{[0.001, 0.999]}(5, 5)$$

$$\mu^{\rho}_{G_i} \sim Beta_{[0.001, 0.999]}(5, 5)$$

To adhere to JAGS syntax, the precision parameter λ was used instead of variance, with the same prior set for all parameters:

$$\lambda \sim Gamma(0.5, 0.5)$$

2.3.2 Parameter Recovery

To test model quality in parameter recovery on group level and elucidate whether the model is able to accurately predict parameter values from sparse data, the model was run through 100 iterations of fitting on simulated data, including 12 groups in each iteration ((13+11)/2).

For each iteration, means and standard deviations were sampled from uniform distributions as follows:

$$\mu^{\alpha} \sim Uniform(0, 20)$$

$$\mu^{\omega} \sim Uniform(0.001, 0.999)$$

$$\mu^{\rho} \sim Uniform(0.001, 0.999)$$

With the sigma prior being the same for all parameters:

$$\sigma \sim Uniform(0, 0.1)$$

[redacted]

2.3.3 Parameter Estimation & Group Comparison

For each group type, a hierarchical CC model was fitted with the same configurations as the individual model (40000 iterations, 8000 burnin, 3 chains). Model fit was assessed through inspection of convergence diagnostics, including trace plots and potential scale reduction factor (PSRF = $\hat{R}$) values. For the resulting posteriors for the inferred μ's for each trait ρ, α, and ω, a .95 credibility interval was calculated and the result plotted for each of the two strategy group. The overlap coefficient for the two posteriors was calculated using the *bayestestR* package (R Core Team, 2021). This calculates the area under the curve for the common area of two posteriors and returns a percentage of the total area under the curve.

3. Results

3.1 CC Model Validation

3.1.1 Parameter Recovery [redacted]

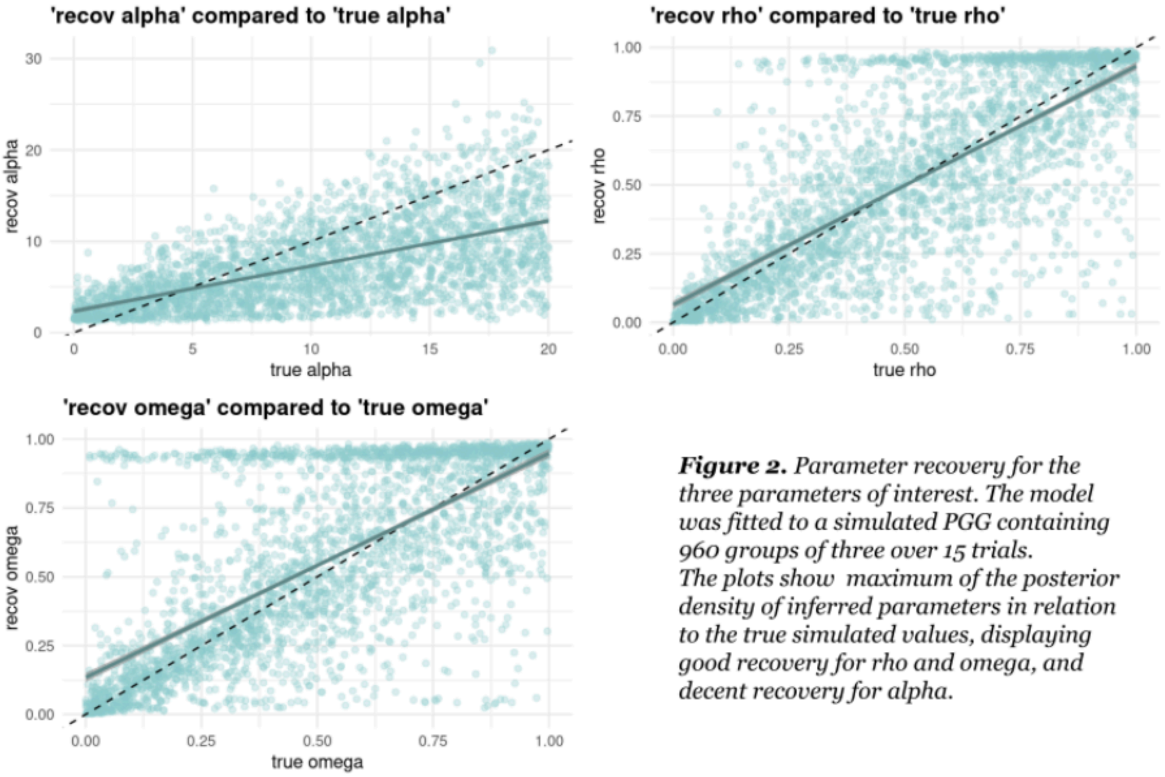


Figure 2. *Parameter recovery for the three parameters of interest. The model was fitted to a simulated PGG containing 960 groups of three over 15 trials. The plots show maximum of the posterior density of inferred parameters in relation to the true simulated values, displaying good recovery for rho and omega, and decent recovery for alpha.*

[redacted]

Figure 2 shows parameter recovery for the CC-model on individual level. Based on visual inspection, the model displays good recovery for both ρ and ω, while it tends to undershoot the estimation of α. Potential causes of this result are further elaborated in the Discussion. Additionally, plots of the inferred parameters as a function of other inferred parameters and other true parameters did not reveal any dependencies between neither of the parameters, see *Appendix A*.

3.1.2 Parameter Estimation [redacted]

3.1.2.1 Convergence Diagnostics

Based on a multivariate PSRF of 1.08 and the majority of parameters having a univariate PSRF ($\hat{R}$) below 1.1, the chains in the model have likely converged. Only parameters with $\hat{R}$ above 1.1 were rho[1,6] ($\hat{R}$=1.25) and rho[2,61] ($\hat{R}$=1.27). Magnitude of the autocorrelation coefficients for lag 1 were inspected visually across all three chains separately, showing only four points of extreme autocorrelation using a threshold of ±0.2. The mean and median autocorrelation in every chain for each parameter was 0, and the min./max values were [-0.67,0.69].

Range of ESS values was broad ranging from a min. of 266.7 to a max of 57476, with a median of 13459.6, indicating half of the parameters have mixed better than what this value indicates, and a 3rd quartile of 25807.9. The range and median suggests that most parameters in the model have reasonable sampling, but with some points of worry, considering the min. ESS relative to the 32000 post-burn-in iterations for each chain.

The parameters with low ESS indicate poor mixing and perhaps high autocorrelation. Trace plots were inspected for exploration of posterior space indicating good mixing overall. Due to having parameters for each individual, trace plots are included herein only for the parameters with the top-three highest PSRF for each parameter of interest (ρ, ω, and α), see *Figure 3*.

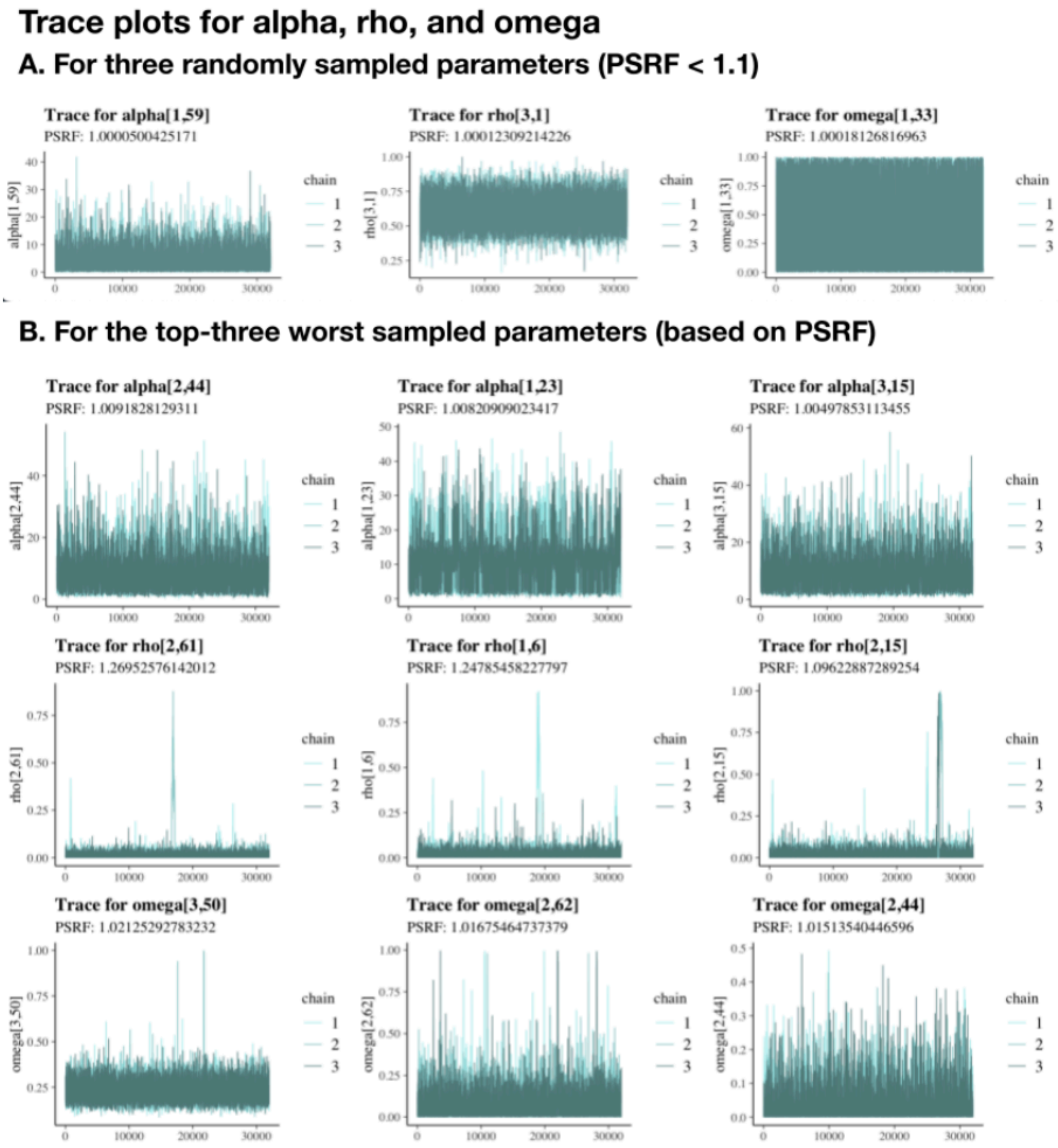
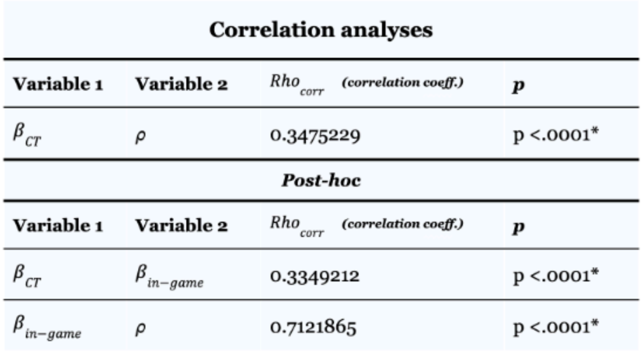


Figure 3: ***A.:*** *Trace plots for three randomly sampled parameters with PSRF < 1.1, showing overall "hairy caterpillars" which indicate good mixing.* ***B.:*** *Trace plots for the top-three parameters with the worst (i.e. highest PSRF) values, showing irregular patterns in mixing, indicating exploration for these parameters was strained and convergence likely not achieved.*

[redacted]

A.



Correlation analyses			
Variable 1	Variable 2	Rho_{corr} (correlation coeff.)	p
β_{CT}	ρ	0.3475229	p <.0001*
Post-hoc			
Variable 1	Variable 2	Rho_{corr} (correlation coeff.)	p
β_{CT}	$\beta_{in-game}$	0.3349212	p <.0001*
$\beta_{in-game}$	ρ	0.7121865	p <.0001*

B.

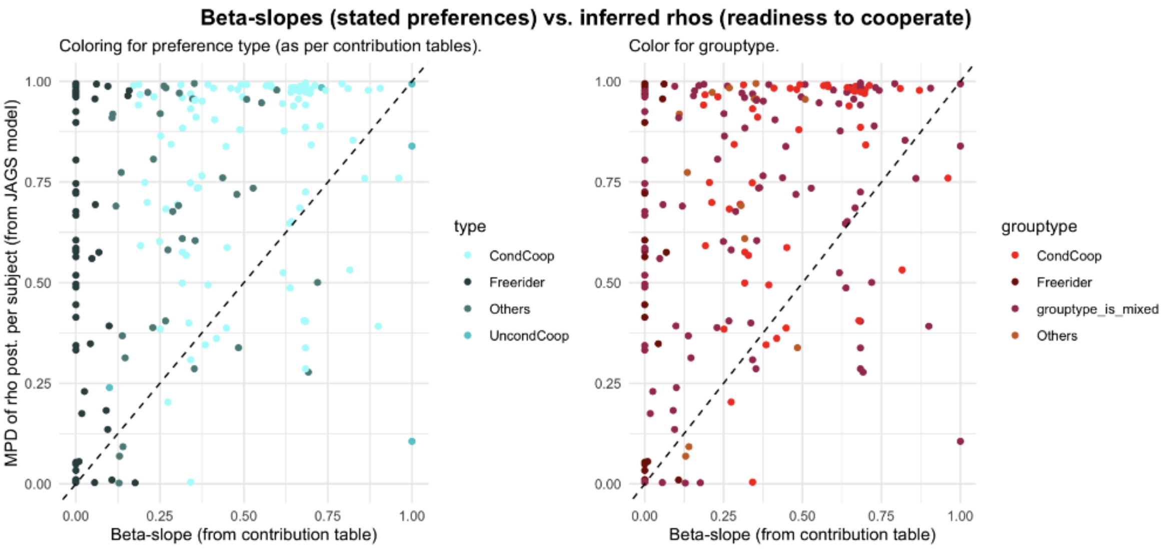


***Figure 4. A**: Table showing results from correlation analyses, both for main and post-hoc. Notably, beta-slopes were normalised between 0-1. **B**: Scatterplot with diagonal line for stated preferences (as per CTs) against inferred rhos (as per the CC model). Two different colorings are made, showing player types and group types. The plot shows the discrepancy between the types derived from the CTs versus the inferred rhos; e.g., a player classified as per the CT as a freerider (visible on the plot as a dark dot to the right) according to the CC model can have a high readiness to cooperate (rho).*

3.1.3 Correlation Analysis [redacted]

A positive significant correlation between normalised βs and inferred ρs was found (Rho_{corr} = 0.35, p < 0.0001*), suggesting a significant association between β-slopes derived from CTs and inferred ϱs from the CC model, see *Figure 4*.

[redacted]

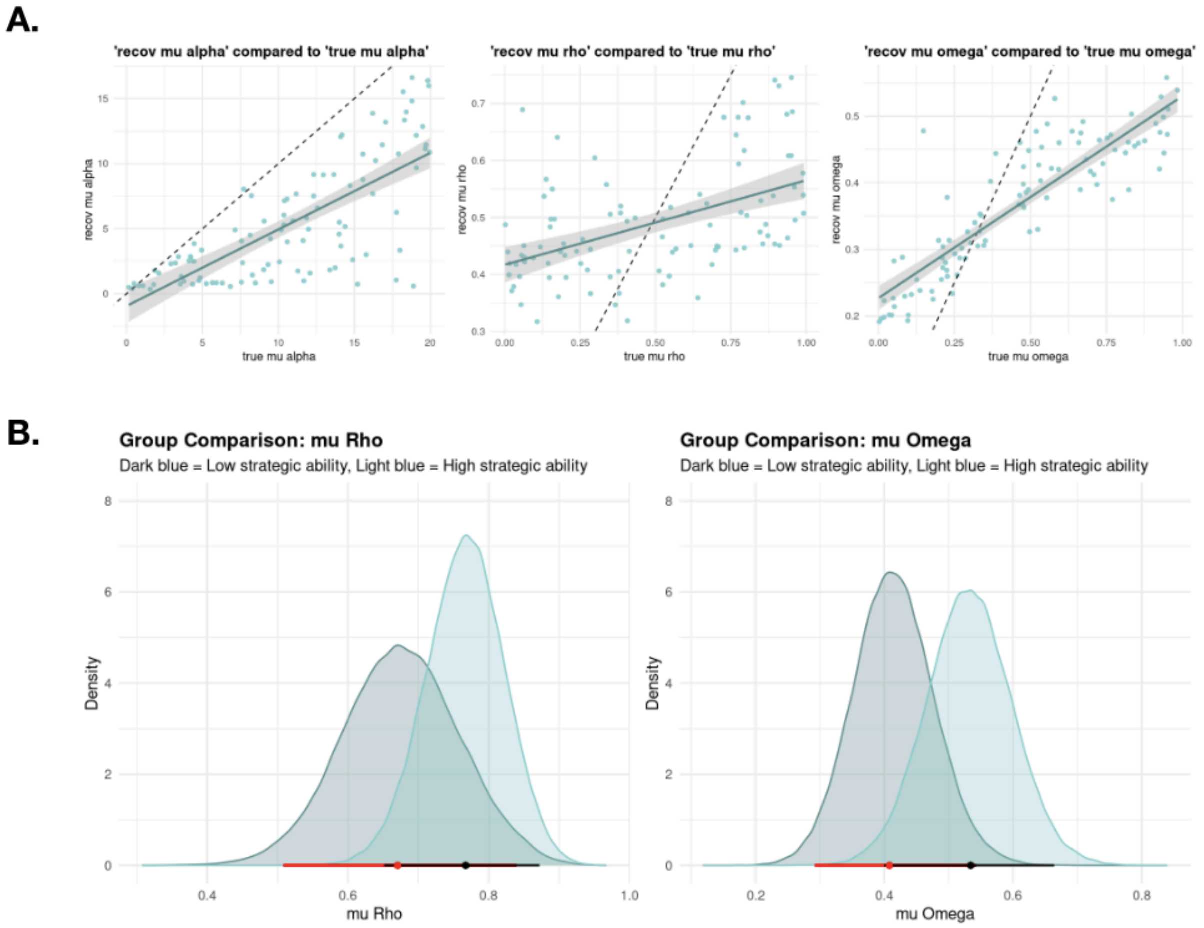


Figure 5. *Hierarchical group comparison.* ***A****: Recovery plots indicate that the model has trouble accurately predicting the means of the underlying group distributions from through 100 iterations containing simulated data of 12 groups of three in a PGG.*
B*: Pairwise plots of the distributions of mu for respectively rho and omega. Though the posteriors seem different, CIs show large overlaps*

3.2 Group Analysis [redacted]

3.2.1 Parameter recovery

Figure 5.A shows parameter recovery for $\mu\alpha$, $\mu\omega$, and $\mu\varrho$, plotted against their true sampled parameter values. It is clear from visual inspection that the model overshoots estimations of smaller values of ϱ and ω, and undershoots for larger values. Furthermore, the model failed to accurately predict λ in all instances, see *Appendix A* for plots. The implications of these results are addressed in the Discussion.

[redacted]

3.2.2 Parameter estimation

Figure 5.B shows posteriors for each of parameter of interest, respectively μρ and μω in the group comparison. In both cases, the mean of the group distribution for high strategy individuals seem to on average be higher, however, the .95 credibility intervals (CI) overlap for both parameters of interest, (μρ: $CI_{G_{high}}$ $[0.653,\ 0.871]$, $CI_{G_{low}}$ $[0.510,\ 0.837]$, μω: $CI_{G_{high}}$ $[0.401,\ 0.663]$, $CI_{G_{low}}$ $[0.295,\ 0.540]$), with the overlap coefficient being respectively 0.50 for μρ and 0.36 for μω. Together with visual inspection of the plotted posteriors, it suggests small but insignificant difference between groups for both μρ and μω (See *Discussion*).

Convergence Diagnostics

Models for the two groups had parameters with varying PSRF and ESS values, indicating that for some parameters, convergence was not achieved. The model for 'strategy-high' showed overall better PSRF values below the threshold of 1.1 for all parameters, whereas the parameters in the 'strategy-low' model for α ($\lambda\alpha$ and μα) and λ for ρ exceeded 1.1. The parameters for α generally showed inefficient sampling having low ESS below 1000 (but notably, this parameter was not a part of the hypothesis set), in contrast to the high ESS for parameter $\lambda\omega$ of >6000. While PSRF for the high-strategy model looked satisfactory to indicate convergence, ESS had variance, the samplings for α being the lowest like for low-strategy. This was also reflected in the trace plots. Autocorrelation also indicated that α was the most problematic parameter, however no values for autocorrelation was above ±0.6.

4. Discussion

[redacted] The present study set out to validate the CC model by correlating the parameter ϱ (readiness to cooperate) with β-slopes as derived from CTs. The study also presented an exploratory analysis assessing group differences in posteriors for parameters ρ and ω (attention to others). The sections below interpret the results taking into account model quality along with discussing implications of results and major limitations tied to the present study.

[redacted]

4.1 Interpretation of Results

4.1.1 H1: CC Model Validation [redacted]

Support was found for H1, although the correlation was weak to moderate: a significant correlation of Rho_{corr} = ~0.35 was found between β_{CT} and the inferred ρs from the CC model for each subject. This means that there is some relationship between the explicit preferences for contributions relative to average contribution of other group members and readiness to cooperate as per the ϱs inferred in the CC model.

Parameter recovery for the model was deemed satisfactory for the parameters ϱ and ω, and less, but still satisfactory for the parameter α, indicating that while imperfect, the estimates are reliable, given that we accept the premise of the CC model as a formalisation of a true decision making process.

According to the multi- and univariate PSRF values, convergence was also achieved, with only two parameters having PSRFs > 1.1. However, the wide range of ESS showed that some parameters had much less efficient sampling than others. Overall, while good parameter recovery and convergence does not guarantee good predictive performance, it does afford a stronger level of confidence for estimates for ϱ and ω, and in turn for the use of ρ in the correlation analysis.

Post-hoc correlations were conducted to investigate the relationships between explicit (derived from CTs) and implicit preferences (derived from in-game behaviour). First, to compare the two using the same scheme of using linear models to derive slopes for preferences, a correlation test was done for β_{CT} and $\beta_{in-game}$. A significant correlation was found between the slopes (Rho_{corr} = ~0.34).

This positive correlation, while significant, is as for the main correlation also only weak to moderate. The reason for it not being higher can naturally stem from many sources, including but not being limited to i) linear models not being the best way to derive slopes, ii) small datasets, iii) difference between explicit and implicit preferences, iiii) manipulation in the Grandjean et al., (2022) study.

Another post-hoc correlation was conducted between $\beta_{in-game}$ and inferred ρs from the CC model, finding a significant and high correlation (Rho_{corr} = ~0.71). The two variables, while derived by different means and with different dependencies and levels of complexity, are derived from the same data; in that sense, the correlation being high is unsurprising, but nonetheless a sanity check.

It is possible that the parameters ω and α would explain the remaining variance, but this is not tested herein and would need more research focusing on finding ways of deriving slopes which better fit the data distribution; the $\beta_{in-game}$ are derived via a highly simplified route of fitting lm() models to subject-subsets of not always normally distributed data, and the intercepts were constrained to be 0 also in cases where this changed the estimate dramatically.

4.1.2 H2 & H3 : Exploratory Group Comparison [redacted]

Little to no support was found for either H2 or H3, respectively hypothesising a higher mean for ρ and ω in high- compared to low-strategy groups. The CIs for the distributions for $\mu\omega$ and $\mu\rho$ for respectively low- and high-strategy groups' overlapped in both cases, indicating no substantial difference between posteriors (assessed pairwise for ω and ρ). These results are further substantiated by the overlap coefficients, which in both cases indicate overlap of posteriors.

However, inspection of the posteriors do hint to some directions, which is in line with the initial hypotheses and theory: From Figure 5.B, there is a tendency towards higher parameter estimates for the μ of the assumed underlying group distribution for high-strategy players compared to low-strategy for both $\mu\varrho$ and $\mu\omega$. For $\mu\rho$ this difference would suggest that individuals in high-strategy groups have a higher readiness to cooperate than those in low-strategy groups. This is perhaps because they see the potential benefit of working together to maximise shared goods.

For $\mu\omega$, which is concerned with the attention to others' behaviour, this would suggest that strategic individuals may be more attentive to what others do to better gauge the most strategic choice. Evidently, it is not a good strategic decision to keep contributing 20 if the other group members are complete freeriders.

The results should be interpreted in the light of the parameter recovery, which demonstrates that the model has trouble accurately predicting both low and high values of $\mu\rho$ and $\mu\omega$. This can potentially skew the result to either side due to stochasticity, which presents an issue both for reliability of the result and validity of the hierarchical model. Furthermore, issues with model fit and lack of strong results may also be due to data scarcity, with only 11 low- and 13 high-strategy groups.

4.2 Implications of Results

[redacted] The results of the correlation tests suggest that there is some disharmony between preferences for conditional cooperation inferred from data and preferences inferred from the contribution tables (CTs). There are several ways to interpret the implications of these results. Yet, all underline the conclusion that it is important to be aware of differences between the two when outlining one's research question.

In Grandjean et al. (2022), CTs are used to classify individuals into player types in order to group them. Subsequently, these inferred player types are used to investigate influences of preferences on contributions. The authors thus demonstrate a common use of CTs as a direct proxy for preferences. However, as Grandjean et al. (2022) also mention, while the practice of gleaning preferences and classifying player types based on the CTs is widely used in the field, "[...] *it is important to consider the possibility that this inference* [of preferences from CTs] *is not always valid.*" (Grandjean et al., 2022, section 4. "Concluding remarks").

The reason may be found in several fundamental differences between the setup of the CT and the PGG, which may affect the relationship: In the CT, an individual matches their contribution 'forward', as they decide their contribution based on *knowledge* about what others contribute. In a PGG, the individual matches 'backwards', as their contribution relies on their *belief* about what others will contribute in the coming round based on evidence from previous trials.

Additionally, the two represent respectively an *explicit* and *implicit* expression of preference for cooperation. The values of the CTs are players' own perceptions of contributions that they *would* make, were they in *some* group where the members on average contributed *x* tokens. In contrast, preferences in a PGG are not explicitly expressed, but instead lie implicitly as drivers of behaviour as modelled in the CC model.

This mismatch can potentially influence the usefulness of the CT. It is entirely possible that people are not adept at perceiving what they would or would not give in a hypothetical scenario. The perception of one's preferred contributions could differ from the 'true' preference type, provided *one* type exists. With the abovementioned caveats regarding deriving β from linear models, this might also partly explain why the correlation between

β_{CT} and ρ was so low; if CTs are inaccurate pictures of people's preferences, their in-game behaviour will naturally differ from the explicit preferences.

[redacted] On the other hand, one could assume the CTs to be accurate, provided that there is such a thing as a permanent behavioural trait for cooperation preferences. Taking this standing point in contrast to the former, the weak correlation could theoretically be the result of e.g. the social manipulation of being in a social game compared to filling out a table.

Qualitative inspections were made into the data, of which subject #7 is an interesting case of example: Subject #7's explicit preferences denoted them as a perfect freerider (having consistently put zeros throughout the CT), while their actual in-game behaviour would have classified them as nearly perfect conditional cooperator ($\rho_7 = 0.99$). While subject #7 naturally only represents a single case, this exemplifies the potential distance between explicit and implicit preferences in a PGG. This is important to consider in contexts where CTs are used as the point against which to compare one's result. Furthermore, it speaks to the use of cognitive models such as the CC-model, which retrieves traits directly from the data instead of from separate tasks.

The differences between CTs and in-game behaviour described above does not mean that contribution tables do not hold value. Instead, they may fuel research questions addressing the impact of e.g. differences between

a) explicit and implicit preferences
b) *knowing* what others give a priori (CTs) versus what you *believe* and/or *hope* the others will give (PGG) (as also stated in Fischbacher & Gächter (2010))

One could also address the impact of a social setting, i.e., what difference the social element makes for your behaviour. All these issues can, notably, also be addressed using the CC model as an interaction between the latent variables, possibly in combination with the CT, provided the information the CT actually holds is relevant for the present research question. As such, both the CC model and the CTs provide interesting cases for future research in a cognitive modelling context for PGGs.

4.3 Limitations and Future Research

4.3.1 Deriving β Using Linear Regression [redacted]

Several choices were made during the extraction of β-slopes for the correlation analyses that potentially influence validity of the estimated coefficients.

Forcing the intercept to be zero assumes that when the average contribution of others is 0, so is one's own contribution (being the stated (explicitly preferred) contribution, and actual (implicitly preferred) contribution for in-game slopes). Furthermore, th β-estimates were normalised to be between 0 and 1 to match the range of ϱ. This essentially means that we only allow for matching and undermatching, and this assumption is quite strong and as is the case for the abovementioned potential "fake freerider" Subject #7, not aligning with all cases. In sum, the simple regression with intercepts through the origin may not capture relationships accurately. However, this is deemed a minor issue for the present data, since results of post-hoc correlation analyses showed similar results to the main analysis when β-slopes were derived from models with an intercept. Furthermore, fitting individual models i) risks overfitting and the predictive power of these models is therefore likely low, further seeding doubt about validity of estimates. Further, ii) we risk having more extreme values than we would have gotten, had we e.g. sampled from a group mean instead. This could have been done by fitting a hierarchical model sampling the subject-specific β-slopes from a group mean, and this may have led to more generalisable estimates.

Other ways to validate the correlations would be to look into nonlinear alternatives of deriving estimates analogous to slopes used in the present analyses, since the data for some subjects indicated nonlinear relationships. Notably, the used CC model herein is a nonlinear model, updating variables dynamically based on priors and impact of new information.

Data for the linear models were found to often be non-normal, meaning the assumption of normality was often broken. This is usually not critical for larger samples given the central limit theorem, but for small samples like this (of respectively 21 contributions in the CT, and 15 in the PGG), non-normality can influence the validity of the confidence intervals around the estimates. Furthermore, a substantial part of the linear models had a presence of heteroscedastic residuals (mostly stemming from the forced origin-intercept) which also can influence reliability of the estimates.

4.3.2 Model Fitting For Validatio [redacted]

The highest PSRF values for the individual CC model were all for the ϱ parameters. The wide range of ESS showed that some parameters had low sampling, indicating that some parameters have not converged well or some parameters have too high autocorrelation of which four were found to exceed ±0.2. Iterations were increased in attempts to improve the sampling, also varying burn-in to comprise 10% of total iterations instead of 20%, but with no substantial improvement. Increases in iterations were stopped at 40000 for computational reasons. This tradeoff between model fit and use of resources is a common one in computational modelling.

For this particular investigation, where ϱ was the main focus, the model configuration used sufficed. However, for parameters that proved difficult to recover, it may be profitable to inspect the set priors to see if sampling efficiency could be improved this way. One such parameter is α which is inherently difficult to estimate as it is only used once for every person: On the first trial, before any evidence about others' behaviour can be taken into account.

4.3.3 Limitations of Cognitive Modelling [redacted]

The difficulty of estimating certain parameters such as α, and even more so for λ, illustrates some of the limitations that computational models of cognition face. Besides the fact that some variables will as per design be difficult to estimate, cognitive models are, as all models, sensitive to data scarcity. This makes it difficult to get reliable estimates and could potentially explain why the model has trouble recovering parameters on group level; it needs significantly more data to accurately estimate a distribution. This limitation becomes relevant in most PGG studies, as lack of data is a general issue in all studies that need participants to conduct an experiment - in this case, one that often only has 10-15 trials.

Additionally, it is important to consider that cognitive models are reductionist. They are formalised versions of complex processes based in theory, and it thus matters greatly which theory they aim to adhere to. The present study focuses solely on the belief-based social learning branch within PGG research, but other theoretical frameworks also exist, e.g., within the school of reinforcement learning (Camerer & Hua Ho, 1999). Together with the fact that cognitive models rely on latent variables which can be difficult to directly assess, it can be difficult to know whether the model holds enough ecological validity to serve as an accurate model of real-world psychological phenomena. Nevertheless, cognitive models present an interesting and useful way to address behaviour in a PGG from a different

[redacted]

perspective than traditional statistical modelling. The latter are often bound (as could be seen in the linear model fitting herein) by just as many assumptions.

4.3.4 Conditional Cooperation Models and Ecological Validity [redacted]

The motivation for the present study lies in the assumption that PGGs are useful for elucidating real human decision making behaviours at all. However, as all models are reductions of true and complex phenomena, there will always be things less accurately captured.

The claim that CTs and the CC model (and ϱ in particular) inform anything about preferences are essentially both assumptions as well. They rely on the idea that there is such a thing a preference 'trait', i.e. how much you would contribute given your belief about others' contribution. The main difference between the CT and CC derived preference is that for the former, preferences are derived from a self-report measure separate from the actual PGG in many ways. This casts doubt on the utility of it as per the above discussion. In the latter, preferences are derived from actual behaviour using a nonlinear cognitive model informed dynamically with both priors and actual evidence from the game. This, depending on one's research question, may yield a more accurate picture of a person's preferences.

If the point of CTs is still valid for a given research question at hand, it may be profitable to fill out CTs post-gameplay as this may provide a more accurate picture of explicitly stated preferences. Ideally, this should be tested using a within-subject design (in contrast to in (Fischbacher & Gächter, 2010)) to see whether there is a difference between pre-/post-gameplay in awareness of preference pattern. Regardless, it may be profitable to at least include trial rounds of a (potentially simulated game) basic PGG before doing the actual gameplay if one aims to infer preference patterns from this. This would both weed out anomalies in the data stemming from potential misunderstandings regarding game instructions along with possibly providing for better data (unless one is studying more immediate effects).

5. Conclusion [redacted]

The present study sought to validate the cognitive model of conditional cooperation presented by Skewes & Nockur (2023), a formalisation of Fischbacher & Gächter‘s (2010) schema for social learning in public goods games. Focus was on whether the model parameter ρ holds validity as an expression of preference for condition cooperation.

[redacted]

Additionally, the model was modified to be hierarchical, to investigate idiosyncratic relations in the data as per a strategy-level matching of players in groups.

A significant, yet low, positive correlation was found for CC-model-inferred ρs and βs derived from contribution tables (CTs), supporting H1. The low correlation indicates discrepancies between stated preferences in CTs and inferred readiness to cooperate from in-game behaviour from the CC model. Possible reasons for this discrepancy include differences in information gleaned from CTs vs in-game behaviour and in-/deflated estimates as per the constrained origin-intercept in linear models from which βs were derived. The CC model passed parameter recovery and model convergence thresholds satisfactorily but not perfectly with areas of improvement for parameter α in particular.

Little to no support was found for underlying group differences when investigating attention to others (ω) and ρ for high- vs low-strategy matched players in groups. Directions of possible effects to be found in replications with more data are discussed, such as high-strategy matched groups indicating posterior distributions for both parameters with slightly higher means than for low-strategy matched groups. However, the results should be viewed in the light of unstable recovery and model convergence.

Though the CC model holds certain limitations, our results suggests differences between CTs and ϱ which may supports the use of the CC-model as way to model behaviour in PGG, however, the use of CTs, given ones research question, may still be relevant. Future research is needed to further elucidate the reasons for discrepancy between CTs and inferred game-behaviour along with the possible differences in strategy-matched groups.

[redacted]

6. Bibliography

Camerer, C., & Hua Ho, T. (1999). Experience-weighted Attraction Learning in Normal Form Games. *Econometrica*, *67*(4), 827–874. https://doi.org/10.1111/1468-0262.00054

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, *14*(1), 47–83. https://doi.org/10.1007/s10683-010-9257-1

Dong, Y., Zhang, B., & Tao, Y. (2016). The dynamics of human behavior in the public goods game with institutional incentives. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep28809

Egashira, S., Taishido, M., Hands, D. W., & Mäki, U. (Eds.). (2021). *A Genealogy of Self-Interest in Economics*. Springer. https://doi.org/10.1007/978-981-15-9395-6

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. SAGE Publications Ltd. https://uk.sagepub.com/en-gb/eur/discovering-statistics-using-r/book236067

Fischbacher, U., & Gächter, S. (2010). Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments. *American Economic Review*, *100*(1), 541–556. 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. https://doi.org/10.1016/S0165-1765(01)00394-9

Grandjean, G., Lefebvre, M., & Mantovani, M. (2022). Preferences and strategic behavior in public goods games. *Journal of Economic Behavior & Organization*, *195*, 171–185. https://doi.org/10.1016/j.jebo.2022.01.007

Hiatt, L. M., Brooks, C., & Trafton, J. G. (2022). Validating and Refining Cognitive Process Models Using Probabilistic Graphical Models. *Topics in Cognitive Science*, *14*(4), 873–888. https://doi.org/10.1111/tops.12616

Houser, D., & McCabe, K. (2014). Chapter 2—Experimental Economics and Experimental Game Theory. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics (Second Edition)* (pp. 19–34). Academic Press. https://doi.org/10.1016/B978-0-12-416008-8.00002-4

JAGS - Just Another Gibbs Sampler. (n.d.). Retrieved 6 January 2024, from https://mcmc-jags.sourceforge.io/

Jeung, H., Schwieren, C., & Herpertz, S. C. (2016). Rationality and self-interest as economic-exchange strategy in borderline personality disorder: Game theory, social preferences, and interpersonal behavior. *Neuroscience & Biobehavioral Reviews*, *71*, 849–864. https://doi.org/10.1016/j.neubiorev.2016.10.030

Kagel, J. H., & Roth, A. E. (1995). *The Handbook of Experimental Economics*. Princeton University Press.

Katahira, K. (2016). How hierarchical models improve point estimates of model parameters at the individual level. *Journal of Mathematical Psychology*, *73*, 37–58. https://doi.org/10.1016/j.jmp.2016.03.007

Kurzban, R., & Houser, D. (2005). Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *Proceedings of the National Academy of Sciences*, *102*(5), 1803–1807. https://doi.org/10.1073/pnas.0408759102

Larrouy, L., & Lecouteux, G. (2017). Mindreading and endogenous beliefs in games. *Journal of Economic Methodology*, *24*(3), 318–343. https://doi.org/10.1080/1350178X.2017.1335425

Lugovskyy, V., Puzzello, D., Sorensen, A., Walker, J., & Williams, A. (2017). An experimental study of finitely and infinitely repeated linear public goods games. *Games and Economic Behavior*, *102*, 286–302. https://doi.org/10.1016/j.geb.2017.01.004

McElreath, R. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan (1st ed.). Chapman and Hall/CRC*. https://doi.org/10.1201/9781315372495

Muller, L., Sefton, M., Steinberg, R., & Vesterlund, L. (2008). Strategic behavior and learning in repeated voluntary contribution experiments. *Journal of Economic Behavior & Organization*, *67*(3), 782–793. https://doi.org/10.1016/j.jebo.2007.09.001

Otten, K., Frey, U. J., Buskens, V., Przepiorka, W., & Ellemers, N. (2022). Human cooperation in changing groups in a large-scale public goods game. *Nature*

[redacted]

Communications, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-34160-5

Ozono, H., Jin, N., Watabe, M., & Shimizu, K. (2016). Solving the second-order free rider problem in a public goods game: An experiment using a leader support system. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep38349

Pereda, M., Capraro, V., & Sánchez, A. (2019). Group size effects and critical mass in public goods games. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-41988-3

Prezenski, S., Brechmann, A., Wolff, S., & Russwinkel, N. (2017). A Cognitive Modeling Approach to Strategy Formation in Dynamic Decision Making. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01335

R Core Team. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/* [Computer software].

Selten, R. (1967). Die strategiemethode zur erforschung des eingeschr�nkt rationale verhaltens im rahmen eines oligopolexperiments. *Beitr�ge Zur Experimentellen Wirtschaftsforschung*, 136.

Skewes, J. C., & Nockur, L. (2023). National inequality, social capital, and public goods decision-making. *Current Research in Ecological and Social Psychology*, *4*, 100112. https://doi.org/10.1016/j.cresp.2023.100112

Su, Y.-S., & Yajima, M. (2021). *R2jags: Using R to Run 'JAGS'* (0.7-1) [Computer software]. https://cran.r-project.org/web/packages/R2jags/index.html

Thielmann, I., Böhm, R., Ott, M., & Hilbig, B. E. (2021). Economic Games: An Introduction and Guide for Research. *Collabra: Psychology*, *7*(1), 19004. https://doi.org/10.1525/collabra.19004

Thöni, C., & Volk, S. (2018). Conditional cooperation: Review and refinement. *Economics Letters*, *171*, 37–40. https://doi.org/10.1016/j.econlet.2018.06.022

van Staveren, I., Sent, E.-M., & Vyrastekova, J. (2015). Gender Beliefs and Cooperation in a Public Goods game Experiment. *Economics Bulletin*, *35*.

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, *8*, e49547. https://doi.org/10.7554/eLife.49547