Rule Following and Cooperation

Pierce Gately

Student No: 20398296

Word Count: 10,383

Supervisors:

Professor Simon Gächter & Professor Robin Cubitt

1st September 2022

A Dissertation presented for

the degree of MRes Economics

Acknowledgements

I am extremely appreciative of the guidance and support I received from my supervisors Professor Simon Gächter and Professor Robin Cubitt.

Contents

Tables

Figures

Introduction

Rules are a ubiquitous part of life. Rule following is one of the cornerstones of cooperation problems. Implementing rules can be a relatively costless way to induce cooperation by appealing to notions of social or moral norms.[1] However, there is also the potential for rules or social norms to backfire, with violations becoming widespread (Bicchieri et al., 2022; Gächter et al., 2022b). In this experiment we seek to extend previous research by examining whether the purpose of a rule and the social consequences of rule violations impact on the rate of rule following. We combine two experimentally validated paradigms: (i) Public Goods Game(s) (PGG) (e.g., Fehr and Gächter, 2000; Herrmann et al., 2008; Cubitt et al., 2011a,b) and (ii) Rule Following Task(s) (RFT) (e.g., Kimbrough and Vostroknutov, 2016, 2018; Gächter et al., 2022b).

Recent global events have emphasised the role of rules and social norms in everyday life, as governments sought to enshrine cooperation by appealing to rules (e.g., mask mandates during Covid-19). Baum et al. (2004) classify rules as directive statements designed to guide behaviour with respect to different choice alternatives, and assert that rules are a means through which one promotes behaviours which benefit the collective in favour of behaviour that generates short term benefits. Accepting this definition and purpose of rules, the link between rules and cooperation is evident. Rules are an essential component to establishing cooperation on a large scale through their ability to direct and govern behaviour.

The process of rule following is governed by both personal and social motives (Gross and Vostroknutov, 2022). Intrinsic motives to follow a rule may stem from internal personal costs of violating a rule (similar to the psychological costs associated with lying (Abeler et al., 2019)), while extrinsic motives may stem from a desire to conform to social norms or signalling concerns.[2] In this experiment we examine both motives side-by-side with participants either faced with a rule that has a social pur-

[1] The distinction between social norms and moral norms is discussed in Elster (2015) and Schram and Charness (2015). See Enke (2019) for an examination of how moral values influence cooperation in loose and tightly knit kinship societies.

[2] Rule following may also be governed by a desire to maintain ones self- and/or social-image, with the environmental context determining which factor is more motivating. See Bašić and Quercia (2022) for evidence of the role self- and social-image concerns play in decisions to lie.

pose (i.e., the provision of a public good) or a rule that has no purpose (i.e., standard RFT).

Cooperation is at the epicentre of many global issues, such as climate change and the Covid-19 pandemic. Long term goals (e.g., averting climate catastrophe, minimising excess deaths from Covid-19) can only be achieved through collective action in the short and medium term. Games employed in economic experiments try to capture the essence of these global issues by creating scenarios in which collective benefits come at the expense of personal cost, and there may be incentives to free ride off others' efforts.

The PGG framework incorporates these features of social dilemmas and has been applied extensively in behavioural and experimental economics. Similarly, RFT encapsulate the feature that individuals have to make a decision between following a rule at some personal extrinsic cost or violating a rule at some personal intrinsic cost. In both settings individuals evaluate the costs and benefits of their prospective actions.

The experiment presented herein was designed in order to examine the relationship between rule following and cooperation, focusing on several intriguing aspects, such as: (i) do rules influence cooperative behaviour, as measured by contributions to public goods; and (ii) does having a social consequence or purpose increase rule following.

Combining both tasks, our findings corroborate previous research that rules increase cooperative behaviour in PGG, whether the rule is enforced or not. Adding a social consequence and purpose to the rule increases the rate of rule following, though this increase was not significant and may have been driven by indirect factors, such as shifting empirical expectations.

The remainder of the paper is structured as follows. Section 2 presents an overview of existing literature. Section 3 outlines the experimental design implemented, consisting of five experimental treatments. Section 4 sets out the hypotheses to be tested in the experiment. Section 5 presents the experiments results. Section 6 concludes and outlines avenues for future research.

Related Literature

This paper is related to experimental literature on rule following and the provision of public goods, and combines both frameworks to examine the relationship between rule following and cooperation.[3] Several important factors are likely to play a role in determining the link between rule following and cooperation, including signalling and self-image concerns, empirical and normative expectations, material incentives, externalities, and preferences for conditional compliance. The paradigm considered herein cannot disentangle all these effects but rather aims to establish a link between rule following and cooperation. Once this link has been established future research will focus further on explanatory factors and underlying mechanisms.

Kimbrough and Vostroknutov (2016) implemented a RFT to examine whether preferences in experimental research are imported from participants. Specifically, they focused on prosocial preferences and implemented several experimental games (ultimatum game, dictator game, PGG, trust game) alongside their RFT, to examine whether participants' proclivity for rule following was associated with greater prosocial behaviour in subsequent stages of the experiment. Participants completed the RFT in Stage 1 of the experiment, before playing one of the experimental games in Stage 2.

In the RFT participants were initially endowed with €8.00 and were informed that they would lose €0.08 of their endowment for every second they spent on the RFT. Participants were confronted with 5 traffic lights and had to decide how long to spend at each one.[4] Participants were informed that the rule pertaining to this task was to "wait at each stop light until it turns green" (p. 615). Given how the experiment was programmed, participants could earn from this task at most €6.00 (if they never followed the rule) and at least €4.00 (if they always followed the rule). The authors asserted that this paradigm mirrors many real-world scenarios in which people are asked to follow a rule while incurring some personal cost.

Gächter et al. (2022b) adapted the RFT from Kimbrough and Vostroknutov (2016),

[3]The focus of this section is reviewing literature related to rule following as this is the component being embedded into the classical PGG framework. For a review of PGG see Zelmer (2003); Gächter (2007).

[4]See page 615 of Kimbrough and Vostroknutov (2016) for a screenshot of the RFT.

focusing on the case in which there is only one traffic light. The task followed the same procedure in the sense that participants were informed of the 'rule' and endowed with an initial amount that decreased for each second spent on the task. Gächter et al. (2022b) used this task to develop a framework for understanding rule compliance, inspired by interdisciplinary research. The authors developed a function comprising the key elements underpinning individuals' decision to either comply or disobey established rules. The compliance function which frames the focus of the authors three experiments is as follows:

$$C = f\left[O, M, X, n(b^n), e(b^e)\right] \tag{1}$$

where:

Individualistic Motives

- O represents obedience and moral obligations;
- M represents material incentives;
- X represents externalities (concern for others of a violation).

Social Expectations and Conditional Preference for Compliance

- b^n represents normative expectations;
- b^e represents empirical expectations;
- $n(b^n), e(b^e)$ represent conditional preferences for compliance.

Within the context of the experiment considered herein we only explicitly measure empirical expectations. There is scope to extend the framework examined here to consider for example normative expectations by utilising the method proposed in Krupka and Weber (2013) (which was implemented in the case of RFT in Gächter et al. (2022b)). Similarly, one could add questions to examine how participants perceive, and the importance they place on, externalities and their own individual moral obligations.[5]

[5]For instance, in the case of moral obligations one could consider the rule orientation scale,

As in Kimbrough and Vostroknutov (2016), Gächter et al. (2022b) compared individual rule following behaviour in an earlier stage of the experiment with behaviour and beliefs in later stages. The authors asked participants about their normative (what people ought to do) and empirical (what people actually do) expectations of others behaviour in the RFT and demonstrated a clear pattern of conditional compliance among participants (i.e., preferences for compliance increase if others also comply or approve of compliance).

Gächter et al. (2022b) also extended the framework in Kimbrough and Vostroknutov (2016) by examining how participants behaved in the RFT when confronted with the behaviour of other participants. After completing the RFT, participants proceeded to the next stage of the experiment in which they saw a screen which portrayed the behaviour of previous participants in a prior RFT. The authors manipulated participants empirical expectations by presenting them with the behaviour of past participants on the task. Participants completed the RFT for the second time by moving 'simultaneously' with past participants.

Participants saw either 1/2/3/4/5/6 additional dots (which represented previous participants) when undertaking the RFT for the second time, with the same instructions as in the previous stage. The authors findings indicate the stark fact that observing any degree of non-compliance (even only 1 among 6) increased participants own non-compliance, indicating negative influences have a larger impact than positive influences on subsequent behaviour. This pattern was also documented in the case of peer nudging in cooperation and honesty tasks by Isler and Gächter (2022).

Kimbrough and Vostroknutov (2018) introduced a 'bucket RFT' in which participants had to choose how to allocate balls between two buckets.[6] The 'rule' was to "put the balls into the blue bucket" (p. 148). Each ball placed in the blue bucket earned participants £0.05 while each ball placed in the yellow bucket earned participants £0.10. Thus, participants had to balance the personal costs and benefits of breaking the rule. The authors examined the relationship between behaviour in

developed by Fine et al. (2016).

[6]See page 148 of Kimbrough and Vostroknutov (2018) for a screenshot of their bucket RFT.

this bucket RFT and choices in Dictator Games and allocation tasks across participants in five countries (Italy, Netherlands, Turkey, Canada, USA). As in their previous experiment, Kimbrough and Vostroknutov (2018) observed heterogeneity in rule following behaviour across participants, with rule following propensity in Stage 1 positively related to norm-consistent dictator choices in Stage 2.

The bucket RFT introduced by Kimbrough and Vostroknutov (2018) has been used to explore several aspects related to norm compliance. For instance, Schneeberger and Krupka (2021) examined the impact of group identification on norm compliance; Gross and De Dreu (2021) explored the prevalence of collaborative cheating in groups composed of rule-followers and/or rule-violators; Gross et al. (2020) demonstrated how the nature of experimental tasks alters participants desire to be paired with rule-followers and/or rule-violators; Gürdal et al. (2020) studied the link between participants norm-compliance and their desire to join institutions (which differed in their enforcement of rules) in which income was entered into the tax pool and redistributed.

The experiment presented below is in a similar vein to the Institution Choice and Redistribution (ICR) task in Gürdal et al. (2020) in which participants could choose to join one of three redistributive institutions, in which individual income was added to the tax pool and redistributed among all participants, over 20 periods. The institutions had differing stated rules associated with them – the announced tax rate was either 0, 50, or 100%. Gürdal et al. (2020) were interested in exploring the relationship between individual's rule following propensity in the bucket RFT and their choice of and behaviour in redistributive institutions with differing levels and enforcement of rules.

In order to examine the spillover effects of institutions Engl et al. (2021) had participants play two simultaneous PGG across three experimental conditions: 1. PGG with no rule; 2. PGG with exogenous rule; and 3. PGG with endogenous rule. The authors varied whether participants played two simultaneous 'standard' PGG in which they were free to contribute as much as they pleased, or one standard PGG and a PGG in which they were penalised for contributing too little, with 'too little' being determined by either an exogenous (decreed by the experimenter) or endogen-

ous (determined through voting by the group) rule.[7] Participants played these two synchronous PGG for 20 periods and the authors found that there was a positive spillover effect on contributions in the standard PGG framework when participants played one of the variations with rules. Additionally, these spillover effects extended to subsequent one shot PGG without enforcement.

Examining the mechanisms behind these positive spillover effects Engl et al. (2021) emphasised the role of increased expectations of others cooperation coupled with a heightened preference for cooperation. Subsequent trust games revealed that increased cooperativeness did not necessarily transmit to other domains, as participants were not more trusting of each other. The authors observed heterogeneity in spillover effects, with effects more pronounced for those who already exhibited a baseline level of voluntary cooperation and degree of positive reciprocity.

Engl et al. (2021) elicited participants propensity to follow rules using the bucket RFT (Kimbrough and Vostroknutov, 2018) with 30 balls. They found that rule following propensity was not related to spillover effects, with higher rule following propensity being unrelated to rule following in the PGG with institutions. Focusing on the Cognitive Reflection Test (CRT; Frederick, 2005) the authors found suggestive evidence that intuitive thinking is positively linked to spillover effects with lower CRT scores linked to higher contributions in other choices for those in one of the institution conditions.

The experiment proposed below modifies the bucket RFT in Kimbrough and Vostroknutov (2018) to consider the effect on rule following when there is a purpose to the rule and when violating the rule materially harms other participants rather than the experimenter. In previous rule following paradigms (Gächter et al., 2022b; Kimbrough and Vostroknutov, 2016, 2018) rule violations did not generate externalities for other participants as the RFT were non-social.[8] Embedding the

[7] The purpose of the rule was to establish a Minimum Contribution Rate (MCR). If participants contributed below the MCR they received a payoff of zero, while if they contributed at least the MCR their payoff was as in the standard PGG (i.e., Equation (3)). In exogenous rule the MCR was set at 20.

[8] Recall that one of the elements of the compliance function (Equation (1)) in Gächter et al. (2022b) was externalities and a concern for others of a violation. Given the social consequences of rule violations, our proposed framework may generate concern for externalities while previous paradigms may not, given the nature of tasks used. While the tasks implemented in Engl et al.

RFT within a PGG framework, thereby adding a social consequence to rules, allows us to examine the relationship between rule following and cooperation. The treatments discussed below outline the extensions to the bucket RFT in Kimbrough and Vostroknutov (2018).

Experimental Design

The experiment was programmed with LIONESS Lab (Arechar et al., 2018; Giamattei et al., 2020) and run online, with participants recruited on Prolific (https://www.prolific.co/)[27/07/22].[9] Participants were pre-screened according to the following criteria: fluency in English and current residence in the UK, and were paid a participation fee of £1.50 (equivalent to £8.20 per hour) for completing all stages of the experiment.

The experiment took on average 11 minutes (SD=6 minutes) to complete. In addition to the participation fee, participants earned a bonus payment depending on the choices they and their group members made during the experiment. The average bonus payment totalled £1.55, resulting in average participant earnings of £3.05 (inclusive of bonus payment and participation fee).[10] During the experiment participants earnings were denoted in terms of points, which were converted to pounds at the end of the experiment at the following rate: 1 point = £0.20.

Ethics approval was obtained from the Nottingham School of Economics Research Ethics Committee (NSE-REC), and the experiment was pre-registered in AEA RCT Registry. The unique identifying number is: AEARCTR-0009755.

The preregistration outlined that 520 participants would be recruited for the experiment, which translated to 104 participants (with 26 groups per treatment[11]) randomly assigned to one of the five treatments. We achieved this target in one

(2021) and Gürdal et al. (2020) did generate externalities for other participants the primary focus of those papers was not the relationship between rule following and cooperation, but rather the spillover effects of institutions and the relationship between rule following and redistributive choices.

[9]Recent papers by Gupta et al. (2021) and Peer et al. (2017) have examined and compared experimental samples from the Lab, Amazon's Mechanical Turk, Prolific, and CrowdFlower.

[10]This was approximately $3.72 at the time of the experiment.

[11]Participants were grouped post-hoc to form groups of 4 in order to determine their bonus payment.

experimental treatment, however, due to a technical issue usable data was collected from 525 participants and the allocation across treatments was not equal. See Table 1 for the final number of participants in each treatment arm.[12] Participants only took part in one treatment.

Table 1: Number of participants across treatments

Baseline PGG	Baseline RFT	PGG with Rule	PGG with Social Rule	RFT and PGG	
				Order 0	Order 1
104	117	93	99	65	47

† RFT – Rule Following Task, PGG – Public Goods Game
‡ Due to a technical issue for two participants in treatment 5 only their decision in the RFT was recorded so the corresponding numbers for the PGG in treatment 5 are: Order 0 – 64; Order 1– 46
‡‡ Due to a technical issue the number of participants in each treatment arm is not equal

There were five experimental treatments which varied the presence and nature of a rule. The experiment focused on the relationship between rule following and cooperative behaviour in social dilemmas, by adapting the 'bucket task' in Kimbrough and Vostroknutov (2018). Across treatments participants were asked to allocate five tokens between a blue and a yellow bucket. The monetary return from tokens in the yellow bucket was larger than from the blue bucket. In treatments with a rule the rule stated: "The rule is to place the tokens in the blue bucket [because it helps the other group members]". In all experimental treatments the primary outcome of interest was the number of tokens placed in the blue bucket, which represented the rate of rule following and cooperation, respectively.

Figure 1 demonstrates the main experimental task which replicates the bucket RFT in Kimbrough and Vostroknutov (2018). Participants had to decide how to allocate 5 tokens (●) between the blue and yellow buckets in their respective task in each of the different treatments. The differences between treatments arose in the instructions they received about the task (either a RFT or PGG) and whether there was (and the nature of) a rule.[13] Each task in the experiment was one-shot, and completed individually by the participant.

[12]See Table C10 for a demographic breakdown of participants across treatments.

[13]See Appendix A for the differences in instructions and Appendix B for differences in the decision screens participants encountered.

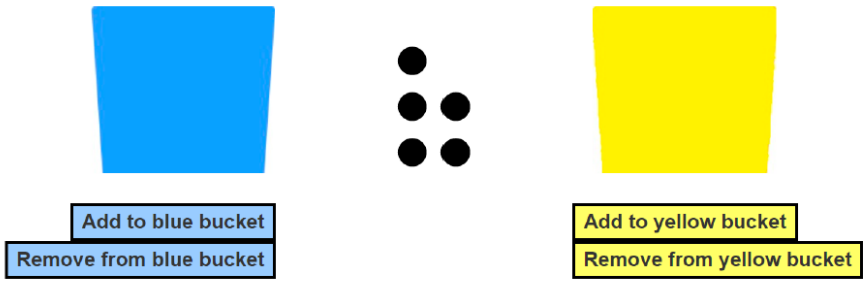

Figure 1: The experimental task

The experimental task was used in all five treatments. Participants had to allocate the 5 tokens between the blue and yellow bucket by clicking the appropriate buttons. Treatments varied whether the blue bucket represented a public good and/or whether a rule stipulating that tokens should be placed in the blue bucket was specified. In each treatment the order of the buckets (i.e., which one appeared on the left/right) was randomised.

The decision task completed in the five treatments is described below. Decisions in this task determined participants bonus earnings. In treatment 5 participants were paid for one randomly chosen task. The treatments are outlined in Table 2. Treatments 1 to 4 were between-subjects. Treatment 5 incorporated treatments 1 and 2 and was within-subjects.

1. **Baseline PGG:** A standard one shot linear PGG in which participants were free to decide how they wished to allocate their 5 tokens between the blue bucket (the public good) and the yellow bucket (tokens kept for oneself). Participants were informed they would form a group of 4, and all contributions to the blue bucket would be doubled to give the value of the public good, which was then distributed equally among group members, irrespective of their own individual contributions. As in standard frameworks we have a marginal per capita return less than 1 and so the prediction is narrowly selfish agents should contribute 0 tokens to the blue bucket. There was no rule in this treatment.

2. **Baseline RFT:** Replicates the bucket RFT in Kimbrough and Vostroknutov (2018). Participants had to decide how to allocate 5 tokens between the blue and yellow bucket. Each token placed in the blue bucket was worth 0.5 points (£0.10), while each token placed in the yellow bucket was worth 1 point (£0.20). Participants saw the following rule when completing the task: *"The rule is to place the tokens in the blue bucket"*.

3. **PGG with Rule:** This task was identical to Baseline PGG except for the

addition of a rule, which stated *"The rule is to place the tokens in the blue bucket"*.

4. **PGG with Social Rule:** This task was identical to Baseline PGG except for the addition of a rule, which stated *"The rule is to place the tokens in the blue bucket because it helps the other group members"*.

5. **RFT and PGG No Rule:** This treatment combined the tasks from Baseline PGG and Baseline RFT with the order of tasks randomised: (i) Order 0: Baseline RFT followed by Baseline PGG; and (ii) Order 1: Baseline PGG followed by Baseline RFT. Participants were paid for one randomly chosen task.

The first two treatments, Baseline PGG and Baseline RFT allowed us to determine a baseline level of cooperative and rule following behaviour, respectively. Combining features from these two treatments allowed us to examine the impact of a social consequence and a purpose on rule following behaviour and the impact of a rule on cooperative behaviour.

Baseline levels of contributions in PGG and rule following in RFT were compared to behaviour in PGG with Rule and PGG with Social Rule to determine whether adding a (social) rule aids in the provision of a public good, and whether having a purpose (i.e., to provide a public good) increases the rate of rule following. RFT and PGG allowed us to examine the relationship between participants' rate of rule following and contributions in PGG.

Table 2: Experimental treatments

	Baseline PGG	Baseline RFT	PGG with Rule	PGG with Social Rule	RFT and PGG	
					Order 0	Order 1
Stage 1	PGG No Stated Rule	RFT No PG	PGG With Stated Rule	PGG Stated Rule refers to "helping other group members"	RFT No PG	PGG No Stated Rule
Stage 2	Empirical Expectations	Empirical Expectations	Empirical Expectations	Empirical Expectations	Empirical Expectations	
Stage 3	×	×	×	×	PGG No Stated Rule	RFT No PG
Stage 4	×	×	×	×	Empirical Expectations	

† RFT – Rule Following Task, PGG – Public Goods Game

* Baseline PGG, Baseline RFT, PGG with Rule, and PGG with Social Rule were between-subjects. RFT and PGG was within-subjects.

‡ Empirical Expectations were elicited via two questions, which varied slightly depending on the task. For PGG the two questions were:1. How many of the **other 3** of your group members placed **all 5** of their tokens in the blue bucket? (*EmpExpAll*); and 2. How many tokens did the **other 3** of your group members **in total** place in the blue bucket? (*EmpExpTotal*). RFT participants were asked to imagine 3 randomly selected Prolific participants and the two questions were: 1. How many of the **3** participants placed **all 5** of their tokens in the blue bucket? (*EmpExpAll*); and 2. How many tokens did the **3** participants **in total** place in the blue bucket? (*EmpExpTotal*). In the within-subjects treatment 'RFT and PGG' participants answered both sets of questions separately, after each respective task.

The structure of the experiment was broadly the same across all treatments. Participants completed the following sequence (with participants in Treatment 5 (RFT and PGG) completing steps (iv)–(vii) twice before moving on to (viii)):

(i) Consent Form

(ii) Prolific ID

(iii) Experiment Information

(iv) Task Instructions

(v) Comprehension Questions related to the task[14]

(vi) Decision Task

(vii) Elicitation of Empirical Expectations[15]

(viii) Cognitive Reflection Test (CRT; Frederick, 2005)

(ix) Exit Questionnaire[16]

In each of the five treatments, participants empirical expectations were elicited after they completed the decision task, via two questions. One question (*EmpExpAll*) asked about how many participants placed *all* of their 5 tokens in the blue bucket, while a second question (*EmpExpTotal*) focused on how many tokens in *total* participants placed in the blue bucket. The two questions were presented as follows after PGG:

EmpExpAll (PGG)

- How many of the **other 3** of your group members placed **all 5** of their tokens in

[14]For PGG there were 4 examples with 2 comprehension questions each, and for RFT there were 2 comprehension questions. Participants were able to return to the instructions at any time when completing the comprehension questions. If they provided an incorrect answer twice they were prompted with the correct answer in order to allow them to continue to the next stage of the experiment.

[15]In treatment 5 participants provided empirical expectations after each task. Participants in treatment 5 were paid for one randomly chosen task and were also paid on the basis of empirical expectations for this randomly chosen task.

[16]All treatments were followed by a questionnaire enquiring about gender, age, political ideology, and whether participants had ever previously participated in a similar study.

the blue bucket?

EmpExpTotal (PGG)

- How many tokens did the **other 3** of your group members **in total** place in the blue bucket?

For empirical expectations related to the RFT participants were asked to imagine 3 randomly selected Prolific participants, and answer the corresponding questions:

EmpExpAll (RFT)

- How many of the **3** participants placed **all 5** of their tokens in the blue bucket?

EmpExpTotal (RFT)

- How many tokens did the **3** participants **in total** place in the blue bucket?

The elicitation of empirical expectations was incentivised in an analogous manner to that employed by Dufwenberg et al. (2011), with participants receiving a 'comparably' large reward (£5.00) if their guesses for both questions were correct, and nothing otherwise. In total across all participants the percentage of consistent empirical expectations (e.g., stating all 3 others would contribute all of their 5 tokens to the blue bucket and then stating the total number of tokens placed in the blue bucket by the 3 others equals 15) was 79.46% (503/633).

The payoff structure in terms of points for participants in the PGG and the RFT were as follows, and reflect the fact that in both settings participants gave up half a point for every token they placed in the blue bucket over the yellow bucket with the PGG offering the potential return of the public good:

$$\pi_i^{\text{RFT}} = \left(\frac{1}{2} \times \text{blue}_i\right) + (\text{yellow}_i - \text{blue}_i) \qquad (i = 1, \ldots, n) \qquad (2)$$
$$\Leftrightarrow -\left(\frac{1}{2} \times \text{blue}_i\right) + \text{yellow}_i$$

$$\pi_i^{\text{PGG}} = \frac{\left(2 \times \sum_j \text{blue}_j\right)}{4} + (\text{yellow}_i - \text{blue}_i) \qquad (j = 1, \ldots, 4) \qquad (3)$$
$$\Leftrightarrow -\left(\frac{1}{2} \times \text{blue}_i\right) + \text{yellow}_i + \frac{\left(2 \times \sum_{j \neq i} \text{blue}_j\right)}{4}$$

where: blue_i and yellow_i are the number of tokens participant i placed in the blue and yellow bucket, respectively.[17]

From (2) and (3) we can see that in PGG for any group member if each of the other 3 group members contribute 0 tokens to the blue bucket, the PGG is payoff equivalent to the RFT, with each token placed in the yellow bucket worth 1 point and each token placed in the blue bucket costing participants 0.5 points, as they forego earning 1 point by placing a token in the yellow bucket.

Participants potential bonus earnings ranged from 2.5 points (£0.50) – 5 points (£1.00) in the RFT, and from 2.5 points (£0.50) – 12.5 points (£2.50) in the PGG.[18] As in previous RFT participants forego a significant proportion of their earnings by adhering to the rule (50 percent of their potential bonus). In the PGG they potentially forego a smaller proportion of their earnings by following the rule (roughly between 20–50 percent depending on the cooperation rate of other group members).[19]

[17]The payoff functions in (2) and (3) only consider material incentives. To more accurately represent how individuals make decisions involving rules these payoff functions can be modified to consider intrinsic motivations or signalling concerns. See Gintis et al. (2001) for an account of signalling concerns in cooperation problems. Kimbrough and Vostroknutov (2016, 2018) consider the role of normative evaluations in individuals payoff functions when considering whether or not to follow a rule.

[18]In addition participants could earn £5.00 if both of their guesses for the empirical expectations questions were correct.

[19]This is the rate of foregone earnings from moving from 0 tokens placed in the blue bucket to 5 tokens placed in the blue bucket. In the case of partial rule following (i.e., 1–4 tokens in the blue bucket) the proportion of foregone earnings compared to 0 tokens in the blue bucket ranges from 4–40 percent, again depending on the cooperation rate of other group members.

Predictions

Based on the experimental design we make the following predictions, with tests reported in the next section. All hypotheses reported below were preregistered in the AEA RCT registry.

Rule following and cooperative behaviour

If people are purely self-regarding they will place all of their tokens in the yellow bucket, because monetary earnings per token are higher than the blue bucket. However, existing evidence suggests that people follow costly rules (e.g., Gächter et al., 2022b; Kimbrough and Vostroknutov, 2016, 2018), and cooperate in one-shot public goods games (e.g., Gächter et al., 2004; Isler et al., 2021). Therefore, the first null hypothesis we expect to reject is that all participants place all of their tokens in the yellow bucket in all treatments (i.e., conform to predictions of self-regarding behaviour). This is a special case of the second null hypothesis that our experimental treatments are unrelated and the fact there is a rule (public good) does not change cooperative (rule following) behaviour.

The experiment is designed so as to allow us examine research questions related to rule following and the social consequence of rules. As previously stated, existing evidence has established that participants are willing to follow a rule even if it is of no social consequence. However, rules are generally established to direct behaviour and serve a social purpose (e.g., rules related to Covid-19 compliance), and thus our experiment aims to establish whether rule following increases if the rule has a social consequence (comparing Baseline RFT and PGG with Rule). Additionally, messaging around rules may emphasise the purpose of the rule (e.g., messaging about masks during Covid-19 which state wearing a mask helps protect yourself and others)[20] and our experiment allows us to test whether rule following increases when the rule has a purpose (comparing PGG with Rule and PGG with Social Rule).

Based on previous findings and research we expect that rule following should be

[20]For example, see Centers for Disease Control and Prevention. (n.d.). Use masks to slow the spread of Covid-19. Centers for Disease Control and Prevention. Retrieved July 18, 2022, from https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/masks.html.

higher when rules have a social consequence and a purpose, and that contributions to the PGG should be higher in the presence of a rule, leading to the following ordered hypotheses:

Hypothesis 1. *The rate of rule following (number of tokens in the blue bucket) should be ordered as follows: Baseline RFT (T2) ≤ PGG with Rule (T3) ≤ PGG with Social Rule (T4).*

Hypothesis 2. *Contributions to the Public Good (number of tokens in the blue bucket) should be ordered as follows: Baseline PGG (T1) ≤ PGG with Rule (T3) ≤ PGG with Social Rule (T4).*

The relationship between rule following and cooperation

Previous research has examined the relationship between individual rule following propensity and behaviour in economic games (e.g., Kimbrough and Vostroknutov, 2016, 2018), and documented a link between individuals' prosociality and rule following proclivity. Treatment 5 (RFT and PGG) is a within-subjects treatment that allows us to replicate previous findings and examine how individual behaviour in both the RFT and the PGG are correlated with each other. We formulated the following hypothesis in relation to participants' behaviour in Treatment 5.

Hypothesis 3. *In the within-subjects treatment RFT and PGG (T5), rule following propensity in the RFT and contributions to the Public good in PGG, as measured by the number of tokens in the blue buckets, should be positively correlated with each other.*

The role of empirical expectations

Individuals' expectations about others' behaviour are an important factor which likely influences their rule following and cooperative behaviour. Previous research has demonstrated individuals' conditional preferences for cooperation (e.g., Fischbacher et al., 2001; Fischbacher and Gächter, 2010) and conforming to a rule (e.g., Gächter et al., 2022b). As such we expect that people's expectations about the number of tokens other people put in the blue bucket (i.e., their empirical expecta-

tions) should be correlated with their own behaviour. In the experiment empirical expectations are elicited, in two ways, as participants' guesses of: (i) the number of other participants who placed *all* of their 5 tokens in the blue bucket (EmpExpAll); and (ii) the *total* number of tokens placed in the blue bucket by other participants (EmpExpTotal). We formulated the following hypothesis.

Hypothesis 4. *Own rule following behaviour, as measured by the number of tokens in the blue bucket, should be positively correlated with participants' empirical expectations of others' allocations to the blue bucket.*

As evidenced by previous research we expect that the presence of a rule should influence people's behaviour (as stipulated by Hypotheses 1–3), and we also expect that behaviour and empirical expectations should be correlated (Hypothesis 4). Combining these hypotheses leads us to expect that empirical expectations should increase in the presence of a stated rule and with the social nature of the rule. We therefore expect that empirical expectations are lowest in the PGG with no rule and highest in the PGG with Social Rule. To test this we formulated the following hypothesis:

Hypothesis 5. *The presence of a rule influences empirical expectations which should be ordered as follows: Baseline PGG (T1)* $\leq$ *Baseline RFT (T2)* $\leq$ *PGG with Rule (T3)* $\leq$ *PGG with Social Rule (T4).*

The role of cognitive reflection

Frederick (2005) examined "cognitive style" (intuitive or reflective), by using the number of correct answers on the three-item CRT as a test score. The main feature of each of the three questions is that they have both a logically correct and an intuitively appealing but incorrect answer. CRT scores range from 0 to 3, representing the number of logically correct answers. Frederick (2005) focused primarily on the relationship between cognitive style and risk and time preferences.[21]

In this study we examine the relationship between cognitive reflection and cooper-

[21] One could possibly conceive of the PGG framework as being related to risk preferences (Jing and Cheo, 2013), given the reliance of payoffs on others' decisions. This may reflect why a substantial proportion of participants are conditional cooperators (Fischbacher et al., 2001; Fischbacher and Gächter, 2010; Cubitt et al., 2017), as they seek to mitigate (financial or emotional) risk by only contributing when others also contribute.

ative/rule following behaviour. If participants answer the CRT questions intuitively they will have a lower CRT score and we expect that engaging in intuitive thinking should increase the rate of rule following, whereas reflective thinkers (those with higher CRT scores) should better understand the incentive to breaking the rules, holding other reasons to follow rules constant. Thus, we formulated the following hypothesis.

Hypothesis 6. *Intuitive thinking influences rule following and scores on the CRT should be negatively correlated with the rate of rule following, as measured by the number of tokens placed in the blue bucket.*

Results and Discussion

The number of tokens participants placed in the blue bucket measured their co-operative or rule following behaviour, and was the primary outcome of interest in the experiment. The proposed framework offers a straightforward way to examine rules with and without social externalities alongside each other, while maintaining the feature that rules are personally costly from previous paradigms. Below we present some summary statistics before turning to tests of the above mentioned hypotheses.[22]

Figure 2 captures the distribution of tokens placed in the blue bucket by participants across the experiment. The percentage of 'complete' rule followers (i.e., those who place all 5 tokens in the blue bucket) was less than in previous studies (e.g. Gächter et al., 2022b) which may be due to the fact here we allow for partial rule following.[23]

In contrast to Kimbrough and Vostroknutov (2018) we did not observe a bimodal distribution, rather a spike at 'complete' rule following when pooling decisions from the RFT and PGG tasks. This trend of a spike at 5 tokens placed in the blue bucket holds when looking across treatments, except for Baseline PGG, and the second task participants completed in the within-subjects treatment RFT and PGG.

[22] Appendix C contains additional results and robustness checks.

[23] See Appendix C (Figures C1–C3) for breakdowns of the distribution by treatment arm and by task in RFT and PGG.

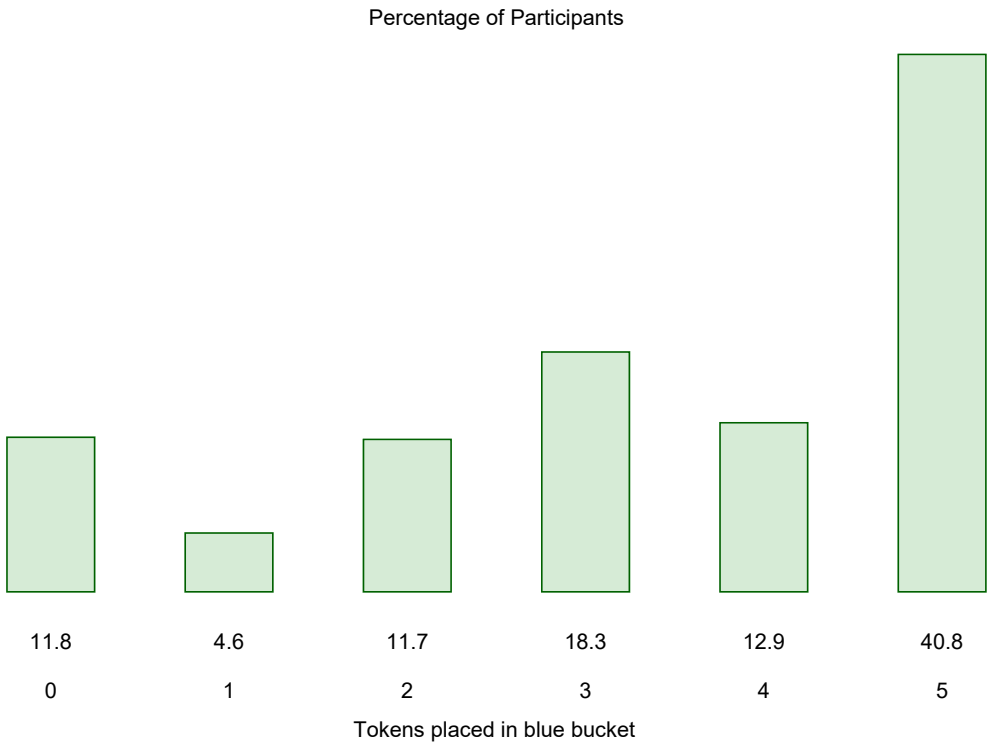

Figure 2: Distribution of tokens placed in the blue bucket for all participants

The above figure shows the percentage of participants across all 5 treatments who respectively placed 0, 1, 2, 3, 4, 5 tokens in the blue bucket. The number below each green bar corresponds to this percentage. In treatment 5 participants decided how to allocate 5 tokens on two separate occasions.

Table 3 presents the mean and median number of tokens placed in the blue bucket across treatments. Comparing the mean number of tokens placed in the blue bucket in Baseline PGG ($M = 3.13$, $SD = 1.43$) and Baseline RFT ($M = 3.47$, $SD = 1.87$) indicates that baseline rule following is higher than baseline cooperative behaviour. Looking at PGG with Rule ($M = 3.72$, $SD = 1.48$) and PGG with Social Rule ($M = 3.97$, $SD = 1.31$) it appears that introducing a rule into the cooperative framework increased contributions to the public good above baseline levels. This potentially suggests that rules are an important driver of cooperative behaviour. The remainder of this section focuses on tests of the above outlined hypotheses.

Table 3: Summary of the number of tokens placed in the blue bucket across treatments

	Baseline PGG	Baseline RFT	PGG with Rule	PGG with Social Rule	RFT and PGG			
					Order 0		Order 1	
					RFT	PGG	RFT	PGG
Mean	3.13	3.47	3.72	3.97	3.66	2.89	2.45	3.04
SD	1.43	1.87	1.48	1.31	1.85	1.94	2.20	1.52
Median	3.00	4.00	4.00	4.00	5.00	3.00	2.00	3.00
N	104	117	93	99	65	64	47	46

† RFT – Rule Following Task, PGG – Public Goods Game, SD – Standard Deviation, N – Number of participants
‡ In treatment RFT and PGG, Order 0 was RFT followed by PGG, Order 1 was PGG followed by RFT
* Due to a technical issue for two participants only their decision in the RFT was recorded in treatment RFT and PGG

Rule following and cooperative behaviour

Rules are established to serve some function and often have a social consequence. For instance, traffic lights direct and manage the flow of traffic. Ignoring the rules codified in traffic lights has negative social consequences, such as increased traffic accidents and injuries. Comparing behaviour in Baseline RFT to behaviour in PGG with Rule and PGG with Social Rule allowed us to determine whether who is affected by a rule violation impacts rule following behaviour.

The ordering proposed by Hypothesis 1 was tested using the non-parametric Jonckheere-Terpstra test for ordered alternatives, and additionally by running the following OLS regression:

$$\text{blue}_i = \alpha_0 + \alpha_1 \text{T3}_i + \alpha_2 \text{T4}_i + \alpha_3 \text{CRT}_i + \alpha_4 X_i + \varepsilon_i \tag{4}$$

with the number of tokens placed in the blue bucket as the dependent variable regressed on treatment dummies (for PGG with Rule (T3) and PGG with Social Rule (T4)), CRT score, and socioeconomic controls (X_i). Baseline RFT (T2) served as the omitted baseline.

The mean number of tokens placed in the blue bucket were ordered as follows 3.47 (Baseline RFT) $<$ 3.72 (PGG with Rule) $<$ 3.97 (PGG with Social Rule), and a Jonckheere-Terpstra Test for Ordered Alternatives indicated that there was no statistically significant trend of higher mean rule following with treatment levels (from T2, T3, to T4), $\text{T}_{\text{JT}} = 2186$, $z = 1.368$, $p = .0857$.[24]

Regression output corresponding to (4) is presented in Table 4 below. Output in column 1 corresponds to a model with only treatment dummies and CRT score as explanatory variables. Column 2 includes demographic controls. Column 3 adds empirical expectations. When controlling only for demographics there is a significant effect ($p < .05$) for T4 (PGG with Social Rule) on the number of tokens placed in the blue bucket. However, when empirical expectations are added to the model

[24]For all Jonckheere-Terpstra tests we report the one-sided ascending ordered alternative as we specified a particular ordering in the preregistered hypotheses. All other tests are two-sided unless otherwise specified. For all results we refer to a significance level of $p < .05$.

this effect is no longer significant. This suggests that T4 absorbed the effect of empirical expectations which are significantly associated with treatment and rule following (i.e., the number of tokens placed in the blue bucket). Failing to control for empirical expectations biases the estimate of treatment effects.

Table 4: OLS Regression Rule following

	Rate of rule following		
	(1)	*(2)*	*(3)*
T3	.247	.256	.124
	(.230)	(.233)	(.196)
T4	.495*	.497*	.179
	(.217)	(.216)	(.179)
CRT	-.023	-.001	-.041
	(.081)	(.087)	(.072)
Constant	3.508***	3.342***	2.315***
	(.201)	(.528)	(.446)
N	309	308	308
Additional covariates	None	Demographics	Demographics + Empirical Expectations

All results are from ordinary least squares regressions of the rate of rule following (as measured by the number of tokens placed in the blue bucket), with T2 as the omitted baseline, in the first column on a dummy variable for T3, a dummy variable for T4, and CRT score. Regressions in column 2 include demographic controls. Regressions in column 3 include demographic controls and empirical expectations. Robust standard errors in parentheses. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05

In order to test whether adding a purpose in addition to a social consequence increased the rate of rule following we tested the magnitude of the coefficients on T3 and T4. In each of the regressions specified in Table 4 a post estimation Wald test indicated that there was no significant difference in magnitude of the coefficient on T3 and T4 ((i) regression 1: $F(1, 305) = 1.49, p = .2235$; (ii) regression 2: $F(1, 294) = 1.32, p = .2523$; and (iii) regression 3: $F(1, 292) = 0.09, p = .7670$).

Result 1. *Social consequences and purpose did not significantly increase the rate of rule following.*[25]

Comparing behaviour in Baseline PGG to PGG with Rule and PGG with Social Rule allowed us to determine the impact of introducing a rule into a PGG framework, and replicate previous findings.

[25]Result 1 should be interpreted tentatively as a larger sample size may allow us to detect smaller effect sizes.

We tested the ordering proposed by Hypothesis 2 using the non-parametric Jonckheere-Terpstra test for ordered alternatives, and also ran the following OLS regression:

$$\text{blue}_i = \alpha_0 + \alpha_1 \text{T3}_i + \alpha_2 \text{T4}_i + \alpha_3 \text{CRT}_i + \alpha_4 X_i + \varepsilon_i \tag{5}$$

with the number of tokens placed in the blue bucket as the dependent variable regressed on treatment dummies (for PGG with Rule (T3) and PGG with Social Rule (T4)), CRT score, and socioeconomic controls (X_i). Baseline PGG (T1) served as the omitted baseline.

The mean number of tokens placed in the blue bucket were ordered as follows 3.13 (Baseline PGG) $<$ 3.72 (PGG with Rule) $<$ 3.97 (PGG with Social Rule). A Jonckheere-Terpstra Test for Ordered Alternatives indicated that there was a statistically significant trend of higher mean contributions to the public good with treatment levels (from T1, T3, to T4), $\text{T}_{\text{JT}} = 6687$, $z = 4.358$, $p = .0000$.

Regression output corresponding to (5) is presented in Table 5 below. Output in column 1 corresponds to a model with only treatment dummies and CRT score as explanatory variables. Column 2 includes demographic controls. Column 3 adds empirical expectations. Controlling for demographics and empirical expectations, there is a significant ($p<.05$) effect of treatment on cooperative behaviour, indicating that introducing a rule into the PGG framework increased cooperative behaviour.

In each of the regressions specified in Table 5 a post estimation Wald test indicated that there was no significant difference in magnitude of the coefficient on T3 and T4 ((i) regression 1: $F(1, 292) = 1.49, p = .2226$; (ii) regression 2: $F(1, 279) = 2.00, p = .1588$; and (iii) regression 3: $F(1, 292) = 0.36, p = .5496$).

Result 2. *Cooperative behaviour increases in the presence of a rule that has social consequences and a purpose.*

Table 5: OLS Regression Cooperative behaviour

	Contributions to public good		
	(1)	*(2)*	*(3)*
T3	.594**	.621**	.406*
	(.208)	(.213)	(.179)
T4	.843***	.916***	.514**
	(.192)	(.191)	(.165)
CRT	-.016	-.011	.006
	(.074)	(.079)	(.063)
Constant	3.150***	2.626***	1.975***
	(.166)	(.480)	(.457)
N	296	293	293
Additional covariates	None	Demographics	Demographics + Empirical Expectations

All results are from ordinary least squares regressions of the rate of contributions to the public good (as measured by the number of tokens placed in the blue bucket), with T1 as the omitted baseline, in the first column on a dummy variable for T3, a dummy variable for T4, and CRT score. Regressions in column 2 include demographic controls. Regressions in column 3 include demographic controls and empirical expectations. Robust standard errors in parentheses. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05

The relationship between rule following and cooperation

The treatment RFT and PGG allows us to examine whether the social element of the PGG impacts on behaviour, by utilising a within-subjects design to compare behaviour in both settings. The order of tasks was randomised to control for possible order effects, and bonus payments were determined by choosing one task at random to control for potential income effects when comparing behaviour across treatments.

Hypothesis 3 was tested by computing the Wilcoxon signed rank test between the number of tokens in the blue bucket in the RFT and the number of tokens in the blue bucket in the PGG, separately for each sequence and for the pooled data disregarding the sequence. In addition, we compared the number of tokens placed in the blue bucket in the between-subjects treatments T1 and T2 using a Mann-Whitney test. We also ran the following OLS regression:

$$\begin{aligned} \text{blue}_i =& \alpha_0 + \alpha_1 \text{PGG Dummy}_i + \alpha_2 \text{Order Dummy}_i \\ &+ \alpha_3 \text{PGG Dummy} \times \text{Order Dummy}_i + \alpha_4 \text{CRT}_i + \alpha_5 X_i + \varepsilon_i \end{aligned} \tag{6}$$

with the number of tokens placed in the blue bucket as the dependent variable

regressed on a PGG Dummy (1 if the blue bucket is the PGG; 0 if it is RFT), an Order Dummy (0 if sequence is RFT then PGG; 1 otherwise) and an interaction variable PGG Dummy × Order Dummy, CRT score, and socioeconomic controls (X_i). Because we observed each participant twice (once in the RFT and once in the PGG), we clustered robust standard errors at the level of the individual.

A Mann-Whitney U test indicated that the mean number of tokens placed in the blue bucket in T1 (Baseline PGG, $M = 3.13$) and T2 (Baseline RFT, $M = 3.47$) are statistically different (z = -2.370, $p = .0177$), suggesting that baseline rule following behaviour is higher than baseline cooperative behaviour.

A Wilcoxon signed rank test indicated that there was a statistically significant difference in the mean number of tokens placed in the blue bucket in the RFT ($M = 3.66$) and PGG ($M = 2.89$) in Order 0 of treatment 5 ($z = 3.089$, $p <.005$). A Wilcoxon signed rank test indicated that there was no statistically significant difference in the mean number of tokens placed in the blue bucket in the RFT ($M = 2.45$) and PGG ($M = 3.04$) in Order 1 of treatment 5 (z = -1.266, $p = .205$).

For the pooled data in treatment 5 a Wilcoxon signed rank test indicated that there was a statistically significant difference in the mean number of tokens placed in the blue bucket in the RFT ($M = 3.15$) and in the PGG ($M = 2.95$) ($z = 2.953$, $p < .005$).

Regression output corresponding to (6) is presented in Table 6 below. Output in column 1 corresponds to a model with only PGG and Order dummies and CRT score as explanatory variables. Column 2 includes demographic controls. Column 3 adds empirical expectations. Controlling for demographics and empirical expectations, there is a significant ($p<.05$) effect of PGG and Order on the number of tokens placed in the blue bucket. This indicates that the rate of rule following was higher than the rate of cooperation in treatment 5 and that the order in which tasks were completed influenced behaviour. In Order 1 the RFT was completed second and so the rate of rule following decreased if the RFT was completed after a cooperative task.

Table 6: OLS Regression Relationship between rule following and cooperation

	Tokens placed in the blue bucket		
	(1)	*(2)*	*(3)*
PGG Dummy	-.770**	-.760**	-.469*
	(.252)	(.257)	(.227)
Order Dummy	-1.240**	-1.202**	-.479*
	(.413)	(.403)	(.232)
PGG × Order	1.367**	1.355**	.456
	(.458)	(.466)	(.328)
CRT	-.042	-.104	.018
	(.133)	(.131)	(.080)
Constant	3.741***	3.638***	1.828***
	(.357)	(.693)	(.526)
N	222	222	220
Additional covariates	None	Demographics	Demographics + Empirical Expectations

Robust standard errors clustered at the level of the individual in parentheses. All results are from ordinary least squares regressions of the number of tokens placed in the blue bucket in T5, in the first column on a dummy variable for whether it was a PGG, a dummy variable for the order in which tasks were completed, an interaction effect between the PGG dummy and the order dummy, and CRT score. Regressions in column 2 include demographic controls, and in column 3 include empirical expectations. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05.

As part of their robustness checks Kimbrough and Vostroknutov (2016) ran a Reverse PGG treatment in which participants played a PGG in Stage 1 of the experiment before completing the RFT in Stage 2. The authors noted that there appeared to be more rule violations in the RFT in the reverse treatment than in the non-reverse treatment in which the RFT was completed first. This may indicate that there were spillover effects from the PGG onto behaviour in the RFT. However, regression analysis indicated that in the case of the framework considered by Kimbrough and Vostroknutov (2016) there was no effect of the reverse treatment on rule following behaviour.[26]

As documented by Tables 6 and 14 there appeared to be an order effect in the within-subjects treatment RFT and PGG which combined the experimental tasks of Baseline PGG and Baseline RFT with the order of tasks randomised. This may be a consequence of the fact that both tasks employed similar decision screens in our paradigm. Participants were potentially 'desensitised' to the task after completing it once already. Thus, they may have deemed it more appropriate to either violate the rule when there was no social consequence attached to it (i.e., in the RFT after the

[26]See Section E.1 in the online appendix for Kimbrough and Vostroknutov (2016) for further information.

PGG) or be less cooperative when there was no implicit rule underpinning behaviour (i.e., in the PGG before the RFT).

Result 3. *Baseline rule following was higher than baseline cooperativeness. The number of tokens allocated to the blue bucket was lower in Order 1 (RFT completed after PGG) of T5 (RFT and PGG).*

The role of empirical expectations

As outlined above, expectations of others' behaviour are likely an important determinant of one's own behaviour. Social norms arise from a desire to guide behaviour, and if they are to be successful require a degree of conformity. The extent to which people conform may determine how 'tight' behaviour is (Gelfand et al., 2011). Gächter et al. (2022b) considered empirical expectations as one of the components of their compliance function. Here we captured empirical expectations through two questions, which were positively correlated with each other.

Hypothesis 4 was tested by computing the Spearman rank order correlation between guesses for the two empirical expectations questions and the number of tokens placed in the blue bucket, and additionally by running the following OLS regression separately for each treatment T1 to T4:

$$\text{blue}_i = \alpha_0 + \alpha_1 \text{EmpExpAll}_i + \alpha_2 \text{EmpExpTotal}_i + \alpha_3 \text{CRT}_i + \alpha_4 X_i + \varepsilon_i \quad (7)$$

with the number of tokens placed in the blue bucket as the dependent variable regressed on guesses for the two empirical expectations questions (EmpExpAll and EmpExpTotal), CRT score, and socioeconomic controls (X_i).

Spearman's rank-order correlation indicated there was a strong positive correlation between the rate of rule following and contributions to the public good and expectations of others rule following behaviour and contributions to the public good, with all correlations positive and significant $p < 0.05$ (see Table 7).

Table 7: Spearman's rank-order correlation - Number of tokens placed in the blue bucket and empirical expectations

	Baseline PGG N = 104			Baseline RFT N = 117		
	Blue	EmpExpAll	EmpExpTotal	Blue	EmpExpAll	EmpExpTotal
Blue	–			–		
EmpExpAll	0.59 ($p < .0001$)	–		0.57 ($p < .0001$)	–	
EmpExpTotal	0.35 ($p = .001$)	0.48 ($p < .0001$)	–	0.54 ($p < .0001$)	0.61 ($p < .0001$)	–
	PGG with Rule N = 93			**PGG with Social Rule N = 99**		
	Blue	EmpExpAll	EmpExpTotal	Blue	EmpExpAll	EmpExpTotal
Blue	–			–		
EmpExpAll	0.60 ($p < .0001$)	–		0.52 ($p < .0001$)	–	
EmpExpTotal	0.47 ($p < .0001$)	0.57 ($p < .0001$)	–	0.37 ($p = .0004$)	0.52 ($p < .0001$)	–

† RFT – Rule Following Task, PGG – Public Goods Game, N – Number of participants
‡ p values in parentheses.

Regression output corresponding to (7) is presented in Table 8 below, separately for each treatment T1 to T4. Output in columns 1 correspond to a model with only EmpExpAll, EmpExpTotal, and CRT score as explanatory variables. Output in columns 2 include demographic controls. Controlling for demographics, there is a significant ($p<.05$) effect of EmpExpAll on participants own allocation of tokens to the blue bucket. EmpExpTotal only appeared to be significantly correlated with own behaviour in Baseline RFT and PGG with Rule.

Table 8: OLS Regression Role of Empirical Expectations

	Tokens placed in the blue bucket							
	(T1 1)	*(T1 2)*	*(T2 1)*	*(T2 2)*	*(T3 1)*	*(T3 2)*	*(T4 1)*	*(T4 2)*
EmpExpAll	.621***	.602***	.588***	.492**	.535***	.544***	.454***	.510***
	(.106)	(.096)	(.159)	(.173)	(.125)	(.136)	(.110)	(.109)
EmpExpTotal	.034	.044	.123***	.138***	.056	.072*	.030	.037
	(.028)	(.026)	(.032)	(.032)	(.031)	(.032)	(.026)	(.025)
CRT	-.066	-.037	-.181	-.196	.078	.113	.033	-.010
	(.091)	(.092)	(.120)	(.130)	(.132)	(.132)	(.096)	(.114)
Constant	1.987***	1.163	1.753***	1.954**	2.079***	2.626**	2.670***	2.542**
	(.267)	(.798)	(.336)	(.727)	(.395)	(.795)	(.329)	(.775)
N	104	102	117	117	93	92	99	99
Additional covariates	None	Demographics	None	Demographics	None	Demographics	None	Demographics

Results from OLS regressions of the number of tokens placed in the blue bucket on empirical expectations (i.e., number of participants who placed all 5 of their tokens in the blue bucket (EmpExpAll) and total number of tokens placed in the blue bucket by other participants (EmpExpTotal)), and CRT score. Column 2 for each treatment include demographic controls. Robust standard errors in parentheses. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05

For T5 Hypothesis 4 was tested by computing the Spearman rank order correlation between guesses for the two empirical expectations questions and the number of tokens placed in the blue bucket in the RFT and PGG respectively, and additionally by running an analogous OLS regression to (7) for T5, that is, separately for RFT and PGG, controlling for order:

$$\text{blue}_i = \alpha_0 + \alpha_1 \text{EmpExpAll}_i + \alpha_2 \text{EmpExpTotal}_i + \alpha_3 \text{Order}_i + \alpha_4 \text{CRT} + \alpha_5 X_i + \varepsilon_i \quad (8)$$

Spearman's rank-order correlation indicated there was a strong positive correlation between participants behaviour and empirical expectations of others behaviour in both tasks of the within-subjects treatment RFT and PGG, with all correlations positive and significant $p < 0.05$ (see Table 9).

Table 9: Spearman's rank-order correlation - Number of tokens placed in the blue bucket and empirical expectations T5

	RFT Order 0 N = 64			RFT Order 1 N = 47		
	Blue	EmpExpAll	EmpExpTotal	Blue	EmpExpAll	EmpExpTotal
Blue	–			–		
EmpExpAll	0.72	–		0.77	–	
	($p < .0001$)			($p < .0001$)		
EmpExpTotal	0.65	0.77	–	0.80	0.65	–
	($p < .001$)	($p < .0001$)		($p < .0001$)	($p < .0001$)	
	PGG Order 0 N = 64			**PGG Order 1 N = 45**		
	Blue	EmpExpAll	EmpExpTotal	Blue	EmpExpAll	EmpExpTotal
Blue	–			–		
EmpExpAll	0.69	–		0.69	–	
	($p < .0001$)			($p < .0001$)		
EmpExpTotal	0.73	0.79	–	0.44	0.48	–
	($p < .0001$)	($p < .0001$)		($p = .0074$)	($p = .0029$)	

† RFT – Rule Following Task, PGG – Public Goods Game, N – Number of participants
‡ p values in parentheses.
* In Order 0 the sequence was RFT followed by PGG. In Order 1 the sequence was PGG followed by RFT.

Regression output corresponding to (8) is presented in Table 10 below, separately for each task (RFT and PGG) in T5. Output in columns 1 correspond to a model with only EmpExpAll, EmpExpTotal, an order dummy, and CRT score as explanatory variables. Columns 2 include demographic controls. Controlling for demographics, there is a significant (p<.05) effect of both EmpExpAll and EmpExpTotal on the number of tokens placed in the blue bucket in both tasks.

Table 10: OLS Regression Role of Empirical Expectations T5

	Tokens placed in the blue bucket			
	(RFT 1)	*(RFT 2)*	*(PGG 1)*	*(PGG 2)*
EmpExpAll	.752***	.761***	.690***	.747***
	(.175)	(.186)	(.114)	(.113)
EmpExpTotal	.145***	.151***	.0956***	.0803**
	(.035)	(.038)	(.026)	(.029)
Order	-.339	-.423	.069	.010
	(.227)	(.225)	(.254)	(.242)
CRTScore	.049	.011	.070	.026
	(.094)	(.095)	(.120)	(.113)
Constant	.843**	1.386*	.865*	1.947**
	(.277)	(.530)	(.359)	(.650)
N	111	111	109	109
Additional covariates	None	Demographics	None	Demographics

Results from OLS regressions of the number of tokens placed in the blue bucket on empirical expectations (i.e., number of participants who placed all 5 of their tokens in the blue bucket (EmpExpAll) and total number of tokens placed in the blue bucket by other participants (EmpExpTotal)), an order dummy, and CRT score. Columns 1 and 2 refer to the RFT with column 2 including demographic controls. Columns 3 and 4 refer to the PGG with column 4 including demographic controls. Robust standard errors in parentheses. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05.

Result 4. *Empirical expectations of others rule following/cooperative behaviour were positively correlated with own rule following/cooperative behaviour. Empirical expectations of the number of participants who were complete rule followers/cooperators were a more important indicator of behaviour than empirical expectations of the total number of tokens placed in the blue bucket.*

Hypothesis 5 was tested by comparing the guesses for the two empirical expectations questions in Baseline PGG vs. Baseline RFT/PGG with Rule/PGG with Social Rule via a Jonckheere-Terpstra test for ordered alternatives, and was tested in RFT and PGG by comparing the guesses for the two empirical expectations questions in PGG vs. RFT tasks via a Wilcoxon signed rank test. Additionally, we ran the following OLS regressions, one for EmpExpAll and one for EmpExpTotal:

$$\text{EmpExp}_i = \alpha_0 + \alpha_1 \text{T2}_i + \alpha_2 \text{T3}_i + \alpha_3 \text{T4}_i + \alpha_4 \text{CRT}_i + \alpha_5 X_i + \varepsilon_i \tag{9}$$

with participants' guesses for the two empirical expectations as dependent variables regressed on treatment dummies (for Baseline RFT (T2), PGG with Rule (T3), and PGG with Social Rule (T4)), CRT score, and socioeconomic controls (X_i). We also ran analogous regressions for T5:

$$\begin{aligned}\text{EmpExp}_i =& \alpha_0 + \alpha_1 \text{PGG Dummy}_i + \alpha_2 \text{Order Dummy}_i \\ &+ \alpha_3 \text{PGG Dummy} \times \text{Order Dummy}_i + \alpha_4 \text{CRT}_i + \alpha_5 X_i + \varepsilon_i\end{aligned} \tag{10}$$

The average empirical expectations of the number of other participants who placed all of their tokens in the blue bucket were ordered as follows 1.55 (Baseline PGG) $<$ 1.70 (Baseline RFT) $<$ 1.83 (PGG with Rule) $<$ 2.09 (PGG with Social Rule). A Jonckheere-Terpstra Test for Ordered Alternatives indicated that there was a statistically significant trend of higher mean empirical expectations of the number of participants who placed all of their tokens in the blue bucket with treatment levels (from T1, T2, T3, to T4), $\text{T}_{\text{JT}} = 9001$, $z = 3.464$, $p = .0003$.

The average empirical expectations of the total number of tokens placed in the blue bucket by other participants were ordered as follows 8.21 (Baseline PGG) $<$ 8.25 (Baseline RFT) $<$ 9.71 (PGG with Rule) $<$ 10.08 (PGG with Social Rule). A Jonckheere-Terpstra Test for Ordered Alternatives indicated that there was a statistically significant trend of higher mean empirical expectations of the number of participants who placed all of their tokens in the blue bucket with treatment levels (from T1, T2, T3, to T4), $\text{T}_{\text{JT}} = 8620$, $z = 3.222$, $p = .0006$.

Regression output corresponding to (9) is presented in Tables 11 and 12 below, separately for EmpExpAll and EmpExpTotal. Output in columns 1 correspond to a model with only treatment dummies and CRT score as explanatory variables. Columns 2 include demographic controls. Controlling for demographics, there is a significant ($p<.05$) effect of T4 (PGG with Social Rule) on both empirical expectations questions. This indicates that appealing to a rule with purpose and a social consequence increases individuals expectations that others will adhere to said rule.

Table 11: OLS Regression Empirical Expectations All

	The role of empirical expectations	
	(EmpExpAll 1)	*(EmpExpAll 2)*
T2	.157	.248
	(.155)	(.154)
T3	.277	.291
	(.172)	(.175)
T4	.536***	.580***
	(.157)	(.156)
CRT	-.054	-.037
	(.051)	(.054)
Constant	1.633***	0.777*
	(.144)	(.360)
N	413	410
Additional covariates	None	Demographics

Results from OLS regressions of empirical expectations (number of participants who placed all of their tokens in the blue bucket (EmpExpAll) and total number of tokens placed in blue bucket by other participants (EmpExpTotal), on Treatment dummies and CRT score. T1 serves as the omitted baseline. Robust standard errors in parentheses. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05

Table 12: OLS Regression Empirical Expectations Total

	The role of empirical expectations	
	(EmpExpTotal 1)	*(EmpExpTotal 2)*
T2	-.001	.142
	(.697)	(.712)
T3	1.526*	1.381
	(.723)	(.748)
T4	1.930**	1.831*
	(.693)	(.713)
CRT	.459*	.392
	(.228)	(.240)
Constant	7.488***	7.596***
	(.624)	(1.606)
N	413	410
Additional covariates	None	Demographics

Results from OLS regressions of empirical expectations (number of participants who placed all of their tokens in the blue bucket (EmpExpAll) and total number of tokens placed in blue bucket by other participants (EmpExpTotal), on Treatment dummies and CRT score. T1 serves as the omitted baseline. Robust standard errors in parentheses. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05

In the regressions specified in Table 11 a post estimation Wald test was run to determine whether there was a significant difference in magnitude of the coefficients on T2, T3, and T4 ((i) regression 1: $F(2, 408) = 3.45, p = .0328$; and (ii) regres-

sion 2: $F(2, 395) = 2.97, p = .0523$). In the regressions specified in Table 12 a post estimation Wald test was run to determine whether there was a significant difference in magnitude of the coefficients on T2, T3, and T4 ((i) regression 1: $F(2, 408) = 4.07, p = .0178$; and (ii) regression 2: $F(2, 395) = 2.82, p = .0609$).

A Wilcoxon signed rank test indicated that there was no statistically significant difference in the mean empirical expectations of the number of other participants who placed all of their tokens in the blue bucket in the RFT ($M = 1.80$) and PGG ($M = 1.52$) in Order 0 of treatment 5 ($z = 1.091$, $p = .275$). A Wilcoxon signed rank test indicated that there was no statistically significant difference in the mean empirical expectations of the total number of tokens placed in the blue bucket by other participants in the RFT ($M = 9.45$) and PGG ($M = 8.84$) in Order 0 of treatment 5 ($z = 0.718$, $p = .473$).

A Wilcoxon signed rank test indicated that there was no statistically significant difference in the mean empirical expectations of the number of other participants who placed all of their tokens in the blue bucket in the RFT ($M = 1.34$) and PGG ($M = 1.76$) in Order 1 of treatment 5 ($z = -1.470$, $p = .142$). A Wilcoxon signed rank test indicated there was a statistically significant difference in the mean empirical expectations of the total number of tokens placed in the blue bucket by other participants in the RFT ($M = 6.02$) and PGG ($M = 8.44$) in Order 1 of treatment 5 ($z = -2.014$, $p = .044$).

For the pooled data in treatment 5 a Wilcoxon signed rank test indicated that there was no statistically significant difference in the mean empirical expectations of the number of other participants who placed all of their tokens in the blue bucket in the RFT ($M = 1.60$) and in the PGG ($M = 1.62$) ($z = 1.786$, $p = .074$). For the pooled data in treatment 5 a Wilcoxon signed rank test indicated that there was no statistically significant difference in the mean empirical expectations of the total number of tokens other participants placed in the blue bucket in the RFT ($M = 8.01$) and in the PGG ($M = 8.67$) ($z = 1.926$, $p = .054$).

Regression output corresponding to (10) is presented in Tables 13 and 14 below, separately for EmpExpAll and EmpExpTotal. Output in columns 1 correspond to a

model with only PGG and Order dummies and CRT score as explanatory variables. Columns 2 include demographic controls. Controlling for demographics, there is a significant ($p<.05$) effect of the interaction terms between PGG and Order on both empirical expectations questions. This indicates that completing the cooperative task first increased empirical expectations of others cooperative behaviour. For EmpExpTotal there was a negative significant ($p<.01$) of the Order dummy suggesting that completing a cooperative task before a RFT decreased empirical expectations of the total number of tokens placed in the blue bucket.

Table 13: OLS Regression Empirical Expectations All T5

	The role of empirical expectations	
	(EmpExpAll 1)	*(EmpExpAll 2)*
PGG Dummy	-.276	-.273
	(.160)	(.163)
Order Dummy	-.552*	-.467
	(.245)	(.240)
PGG × Order	.697*	.701*
	(.295)	(.298)
CRT	-.166	-.170*
	(.087)	(.084)
Constant	2.109***	1.236**
	(.218)	(.442)
N	220	220
Additional covariates	Demographics	Demographics

Robust standard errors clustered at the level of the individual in parentheses. All results are from ordinary least squares regressions of empirical expectations (i.e., number of participants who placed all 5 of their tokens in the blue bucket (EmpExpAll) and total number of tokens placed in the blue bucket by other participants (EmpExpTotal),on a dummy variable for whether it was a PGG, a dummy variable for the order in which tasks were completed, an interaction effect between the PGG dummy and the order dummy, and CRT score. Regressions include demographic controls. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05

Table 14: OLS Regression Empirical Expectations Total T5

	The role of empirical expectations	
	(EmpExpTotal 1)	*(EmpExpTotal 2)*
PGG Dummy	-.604	-.573
	(.716)	(.727)
Order Dummy	-3.352**	-3.018**
	(1.158)	(1.116)
PGG × Order	3.017*	2.996*
	(1.320)	(1.344)
CRT	.123	.032
	(.413)	(.377)
Constant	9.214***	7.576***
	(1.105)	(1.716)
N	222	222
Additional covariates	Demographics	Demographics

Robust standard errors clustered at the level of the individual in parentheses. All results are from ordinary least squares regressions of empirical expectations (i.e., number of participants who placed all 5 of their tokens in the blue bucket (EmpExpAll) and total number of tokens placed in the blue bucket by other participants (EmpExpTotal),on a dummy variable for whether it was a PGG, a dummy variable for the order in which tasks were completed, an interaction effect between the PGG dummy and the order dummy, and CRT score. Regressions include demographic controls. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05

Result 5. *The presence of a rule increases empirical expectations of others rule following/cooperative behaviour. In the within-subjects treatment RFT and PGG, completing the PGG first increased expectations of others cooperative behaviour. Empirical expectations of the total number of tokens placed in the blue bucket were lower in Order 1 (PGG followed by RFT) of T5.*

The role of cognitive reflection

Gächter et al. (2022b) found that including control questions increased the rate of rule violations in their RFT, suggesting that their inclusion may facilitate deliberative thinking which is not conducive to rule following. The potential relationship between deliberative/intuitive thinking and rule following behaviour is formulated and examined in Hypothesis 6, which was tested by computing the Spearman rank order correlation between participants' CRT scores (ranging from 0-3) and rate of rule following behaviour (number of tokens placed in the blue bucket) for each treatment from T1 to T4.

Additionally we hypothesised that the coefficient on CRT would be negative in the regressions specified above, with demographic variables serving as controls and their

analysis solely exploratory. Unlike Kimbrough and Vostroknutov (2016) we did not observe a gender effect on rule following, though this may be a consequence of the unequal gender split in the experiment, which comprised predominantly of female participants ($\approx 64\%$ of participants).[27]

Spearman's rank-order correlation indicated there was no significant correlation between the rate of rule following and contributions to the public good and participants CRT score. Four bilateral Spearman rank order correlations indicated that the correlation between CRT scores and the number of tokens placed in the blue bucket in the four between-subjects treatments ranged from -0.11 to 0.16. According to the conventional significance level of 5% none of these correlations were significant as the p values ranged from .1156 to .8662.[28]

Result 6. *Thinking style (i.e., intuitive/reflective) was not correlated with rule following or cooperative behaviour.*

Exploratory analysis

The within-subjects treatment RFT and PGG generated interesting results which indicated that the order of tasks mattered for rule following behaviour, but not necessarily for cooperative behaviour. To examine whether there were differing spillover effects we compared behaviour in the two respective tasks in T5 (RFT and PGG) to baseline behaviour in T1 (Baseline PGG) and T2 (Baseline RFT).

The mean number of tokens placed in the blue bucket were ordered as follows 2.45 (RFT 2^{nd} in Treatment 5) $<$ 3.47 (Baseline RFT) $<$ 3.66 (RFT 1^{st} in Treatment 5). We did not propose an ordering for these treatments and as such report output from a Kruskal-Wallis test, which indicated that there was a statistically significant difference in the rate of rule following across the three treatments, $\chi^2 = 9.341$, $p = .0094$.

The mean number of tokens placed in the blue bucket were ordered as follows 2.89 (PGG 2^{nd} in Treatment 5) $<$ 3.04 (PGG 1^{st} in Treatment 5) $<$ 3.13 (Baseline PGG).

[27] See Table C10 for the gender breakdown of participants across treatments.

[28] Figure C10 captures the distribution of CRT scores across the 5 experimental treatments, and indicates that the distribution is approximately similar across all treatment arms.

We did not propose an ordering for these treatments and as such report output from a Kruskal-Wallis test, which indicated that there was not a statistically significant difference in the rate of cooperation across the three treatments, $\chi^2 = 0.086$, $p = .9578$.

This suggests that cooperative tendencies may be more innate while rule following tendencies are influenced by contextual factors. Possible spillover effects on rule following may be an artefact of the lack of social consequence or purpose in the RFT in comparison to the PGG.

Result 7 (Exploratory)**.** *Rule following behaviour was influenced by spillover effects in T5 (RFT and PGG), while cooperative behaviour was not impacted.*

Conclusion and Future Research

The experiment proposed herein offers a malleable framework which can be used to analyse the relationship between rule following and cooperation. The design of the experiment allows it to be easily extended in the future to address several interesting questions, some of which are highlighted below.

Utilising a within-subjects design, as in RFT and PGG, allows one to consider the role of potential mechanisms from moral psychology on participants rule following behaviour.[29] For instance one could examine whether concepts such as moral licensing (Blanken et al., 2015), moral cleansing (Brañas-Garza et al., 2013), moral hypocrisy (Batson et al., 1999, 2002), moral disengagement (Bandura et al., 1996; Detert et al., 2008; Moore, 2015) underlie cognitive processes which generate different responses to rules prior to and after a cooperative task.

Bénabou et al. (2020) provide a theoretical and experimental framework with which to elicit moral preferences. Rule following and cooperative behaviour could potentially be incorporated within this framework to examine the influence moral preferences have on such behaviour. Bénabou et al. (2020) focus on an individualistic task

[29]For example, spillover effects evident in previous research may be related to the concept of moral licensing whereby individuals view their moral acts as transactional with positive moral acts offsetting negative moral acts, so that contributing all 5 tokens in one task may allow one to then maintain a positive self-image even if they contribute 0 tokens in the next task.

with participants having to decide whether to forego an amount in order to save a human life (by donating money to a charity). The decision task for participants shares characteristics with PGG as the individual return (either a fixed €100 or between €0–200 depending on condition) is less than the social benefit of donating €350 to charity. Extending the elicitation of moral preferences to tasks which affect more immediate groups may reflect everyday consequences of moral preferences.

One could also explore whether effects persist in the face of feedback or information, for instance by providing participants with information on the normative expectations of other participants, or by playing a repeated PGG with information after each round on the contributions of other group members. This may allow one to focus on interesting questions such as, does peer (Isler and Gächter, 2022) or norm (Bicchieri and Dimant, 2019; Bicchieri et al., 2022) nudging translate from group interactions to individual tasks and behaviour.

Another interesting avenue may be to amend the PGG to explore differences in the maintenance and provision of a public good in the face of a rule.[30] For instance, to consider a maintenance public good framework participants initial decision screen may have all 5 tokens placed in the blue bucket, with participants then allowed to distribute the tokens between the blue and yellow bucket as they see fit.[31] This may also allow for a comparison with literature on (dis)honesty which has revealed that people prefer lies of omission rather than lies of commission (e.g., Schweitzer and Croson, 1999; Mann et al., 2014; Pittarello et al., 2016). In the case of rule following it may similarly be the case that people prefer to passively (rather than actively) violate a rule.

To examine the role of self-image concerns in rule following behaviour one could combine the above paradigm with the framework considered in Falk (2021) in which participants had to make a binary 'moral' choice between an outcome which increased their income but stipulated that another participant would receive a painful (but harmless) electrical shock, or another option in which they did not earn ad-

[30]See Gächter et al. (2022a) for evidence of how the two paradigms are perceived differently by participants in experiments.

[31]Kimbrough and Vostroknutov (2018) explored this aspect of the bucket task in terms of dictator games, altering where tokens were initially placed to create either a 'give' or 'take' frame.

ditional income but no other participant would be shocked. The decision screen manipulated whether participants saw no video, a neutral video, themselves reflected in a mirror, or themselves captured via webcam. This framework could be applied to RFT allowing one to examine whether self-image concerns play a role in driving rule following.

We corroborate previous findings that rules increase cooperative behaviour. Examining further the relationship between rule following and cooperation we document the important role of empirical expectations and potential spillover effects between rule following and cooperative paradigms. We consider a preliminary framework which can be easily expanded over the course of future research to consider key determinants of rule following and cooperation, such as normative expectations, externalities, and moral obligations.

Cooperation is naturally occurring and likely stems from multiple factors, including altruism, reciprocity, and expectations. Central to many contemporary social issues is a desire to foster large scale cooperation. Appealing to social factors can increase individual rule following and cooperative behaviour. As evidenced by the experiments findings the presence of a rule influences how we believe others will behave, which is an important indicator of our own behaviour. The simple stating of a rule may encourage behaviour which benefits future generations.

Understanding the processes that govern and drive large scale cooperation is pivotal to delivering a sustainable future. Examining the relationship between rule following and cooperation we shed new light on the drivers of cooperation, and the relative importance of social consequence and purpose for rule following. Our results suggest interesting avenues for future research to delve further into the determinants of cooperation on a large scale.

References

J. Abeler, D. Nosenzo, and C. Raymond. Preferences for truth-telling. *Econometrica*, 87(4):1115–1153, 2019.

A. Akdeniz and S. Matthew. Normative versus prosocial underpinnings of human decision making. *https://www.otreehub.com/projects/experiment2022til/*.

A. A. Arechar, S. Gächter, and L. Molleman. Conducting interactive experiments online. *Experimental economics*, 21(1):99–131, 2018.

A. Bandura, C. Barbaranelli, G. V. Caprara, and C. Pastorelli. Mechanisms of moral disengagement in the exercise of moral agency. *Journal of personality and social psychology*, 71(2):364, 1996.

Z. Bašić and S. Quercia. The influence of self and social image concerns on lying. *Games and Economic Behavior*, 133:162–169, 2022.

C. D. Batson, E. R. Thompson, G. Seuferling, H. Whitney, and J. A. Strongman. Moral hypocrisy: appearing moral to oneself without being so. *Journal of personality and social psychology*, 77(3):525, 1999.

C. D. Batson, E. R. Thompson, and H. Chen. Moral hypocrisy: addressing some alternatives. *Journal of Personality and Social Psychology*, 83(2):330, 2002.

W. M. Baum, P. J. Richerson, C. M. Efferson, and B. M. Paciotti. Cultural evolution in laboratory microsocieties including traditions of rule giving and rule following. *Evolution and Human Behavior*, 25(5):305–326, 2004.

R. Bénabou, A. Falk, L. Henkel, and J. Tirole. Eliciting moral preferences: Theory and experiment. *Princeton University*, 2020.

C. Bicchieri and E. Dimant. Nudging with care: The risks and benefits of social information. *Public choice*, pages 1–22, 2019.

C. Bicchieri, E. Dimant, S. Gächter, and D. Nosenzo. Social proximity and the erosion of norm compliance. *Games and Economic Behavior*, 132:59–72, 2022.

I. Blanken, N. Van De Ven, and M. Zeelenberg. A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin*, 41(4):540–558, 2015.

P. Brañas-Garza, M. Bucheli, M. P. Espinosa, and T. García-Muñoz. Moral cleansing and moral licenses: experimental evidence. *Economics & Philosophy*, 29(2):199–212, 2013.

G. Charness, A. Samek, and J. van de Ven. What is considered deception in experimental economics? *Experimental Economics*, 25(2):385–412, 2022.

N. J. Cox. Speaking Stata: Multiple bar charts in table form. *The Stata Journal*, 16(2):491–510, 2016.

R. Cubitt, M. Drouvelis, and S. Gächter. Framing and free riding: emotional responses and punishment in social dilemma games. *Experimental Economics*, 14 (2):254–272, 2011a.

R. Cubitt, M. Drouvelis, S. Gächter, and R. Kabalin. Moral judgments in social dilemmas: How bad is free riding? *Journal of Public Economics*, 95(3-4):253–264, 2011b.

R. Cubitt, S. Gächter, and S. Quercia. Conditional cooperation and betrayal aversion. *Journal of Economic Behavior & Organization*, 141:110–121, 2017.

G. d'Adda, M. Drouvelis, and D. Nosenzo. Norm elicitation in within-subject designs: Testing for order effects. *Journal of Behavioral and Experimental Economics*, 62:1–7, 2016.

J. R. Detert, L. K. Treviño, and V. L. Sweitzer. Moral disengagement in ethical decision making: a study of antecedents and outcomes. *Journal of applied psychology*, 93(2):374, 2008.

M. Dufwenberg, S. Gächter, and H. Hennig-Schmidt. The framing of games and the psychology of play. *Games and Economic Behavior*, 73(2):459–478, 2011.

J. Elster. *Explaining social behavior: More nuts and bolts for the social sciences*. Cambridge University Press, 2015.

F. Engl, A. Riedl, and R. Weber. Spillover effects of institutions on cooperative behavior, preferences, and beliefs. *American Economic Journal: Microeconomics*, 13(4):261–99, 2021.

B. Enke. Kinship, cooperation, and the evolution of moral systems. *The Quarterly Journal of Economics*, 134(2):953–1019, 2019.

A. Falk. Facing yourself–a note on self-image. *Journal of Economic Behavior & Organization*, 186:724–734, 2021.

E. Fehr and S. Gächter. Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994, 2000.

A. Fine, B. Van Rooij, Y. Feldman, S. Shalvi, E. Scheper, M. Leib, and E. Cauffman. Rule orientation and behavior: Development and validation of a scale measuring individual acceptance of rule violation. *Psychology, Public Policy, and Law*, 22 (3):314, 2016.

U. Fischbacher and S. Gächter. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American economic review*, 100(1):541–56, 2010.

U. Fischbacher, S. Gächter, and E. Fehr. Are people conditionally cooperative? evidence from a public goods experiment. *Economics letters*, 71(3):397–404, 2001.

S. Frederick. Cognitive reflection and decision making. *Journal of Economic perspectives*, 19(4):25–42, 2005.

S. Gächter. Conditional cooperation: Behavioral regularities from the lab and the field and their policy implications. *In: Frey B.S., Stutzer A., (Eds). Psychology and economics*, MIT Press, 2007.

S. Gächter and J. F. Schulz. Intrinsic honesty and the prevalence of rule violations across societies. *Nature*, 531(7595):496–499, 2016.

S. Gächter, B. Herrmann, and C. Thöni. Trust, voluntary cooperation, and socio-

economic background: survey and experimental evidence. *Journal of Economic Behavior & Organization*, 55(4):505–531, 2004.

S. Gächter, F. Kölle, and S. Quercia. Preferences and Perceptions in Provision and Maintenance Public Goods. *Games and Economic Behavior*, 135:338–355, 2022a.

S. Gächter, L. Molleman, and D. Nosenzo. Why people follow rules. *Working Paper*, 2022b.

M. J. Gelfand, J. L. Raver, L. Nishii, L. M. Leslie, J. Lun, B. C. Lim, L. Duan, A. Almaliach, S. Ang, J. Arnadottir, et al. Differences between tight and loose cultures: A 33-nation study. *science*, 332(6033):1100–1104, 2011.

M. Giamattei, K. S. Yahosseini, S. Gächter, and L. Molleman. Lioness Lab: a free web-based platform for conducting interactive experiments online. *Journal of the Economic Science Association*, 6(1):95–111, 2020.

H. Gintis, E. A. Smith, and S. Bowles. Costly signaling and cooperation. *Journal of theoretical biology*, 213(1):103–119, 2001.

J. Gross and C. K. De Dreu. Rule following mitigates collaborative cheating and facilitates the spreading of honesty within groups. *Personality and Social Psychology Bulletin*, 47(3):395–409, 2021.

J. Gross and A. Vostroknutov. Why do people follow social norms? *Current Opinion in Psychology*, 44:1–6, 2022.

J. Gross, S. Simon, and J. A. Everett. Inferring and enforcing cooperativeness and dishonesty through rule abidance and partner choice. *PsyArXiv*, 2020.

N. Gupta, L. Rigotti, and A. Wilson. The Experimenters' Dilemma: Inferential Preferences over Populations. *arXiv preprint arXiv:2107.05064*, 2021.

M. Y. Gürdal, O. Torul, and A. Vostroknutov. Norm compliance, enforcement, and the survival of redistributive institutions. *Journal of Economic Behavior & Organization*, 178:313–326, 2020.

M. Y. Gürdal, O. Gürerk, Y. Kaçamak, and E. Kart. Conditional commitments and cooperation in public goods: Theory and experimental evidence. 2022.

R. Hakimov and A. Kajackaite. Breaking bad: Malfunctioning institutions erode good behavior. 2022.

B. Herrmann, C. Thöni, and S. Gächter. Antisocial punishment across societies. *Science*, 319(5868):1362–1367, 2008.

O. Isler and S. Gächter. Conforming with peers in honesty and cooperation. *Journal of Economic Behavior & Organization*, 195:75–86, 2022.

O. Isler, S. Gächter, A. J. Maule, and C. Starmer. Contextualised strong reciprocity explains selfless cooperation despite selfish intuitions and weak social heuristics. *Scientific Reports*, 11(1):1–17, 2021.

L. Jing and R. Cheo. House money effects, risk preferences and the public goods game. *Economics Letters*, 120(2):310–313, 2013.

E. O. Kimbrough and A. Vostroknutov. Norms make preferences social. *Journal of the European Economic Association*, 14(3):608–638, 2016.

E. O. Kimbrough and A. Vostroknutov. A portable method of eliciting respect for social norms. *Economics Letters*, 168:147–150, 2018.

E. L. Krupka and R. A. Weber. Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524, 2013.

H. Mann, X. Garcia-Rada, D. Houser, and D. Ariely. Everybody else is doing it: Exploring social transmission of lying behavior. *PloS one*, 9(10):e109591, 2014.

N. Merguei, M. Strobel, and A. Vostroknutov. Moral opportunism as a consequence of decision making under uncertainty. *Journal of Economic Behavior & Organization*, 197:624–642, 2022.

C. Moore. Moral disengagement. *Current Opinion in Psychology*, 6:199–204, 2015.

E. Peer, L. Brandimarte, S. Samat, and A. Acquisti. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70:153–163, 2017.

A. Pittarello, E. Rubaltelli, and D. Motro. Legitimate lies: The relationship between omission, commission, and cheating. *European Journal of Social Psychology*, 46 (4):481–491, 2016.

A. Schneeberger and E. L. Krupka. Determinants of Norm Compliance: Moral Similarity and Group Identification. *Available at SSRN*, 2021.

A. Schram and G. Charness. Inducing social norms in laboratory allocation choices. *Management Science*, 61(7):1531–1546, 2015.

M. E. Schweitzer and R. Croson. Curtailing deception: The impact of direct questions on lies and omissions. *International Journal of Conflict Management*, 1999.

J. Zelmer. Linear public goods experiments: A meta-analysis. *Experimental Economics*, 6(3):299–310, 2003.

A Instructions

Rule Following Task Instructions

Your task

At the start of this task you receive an **endowment** of 5 tokens (● ● ● ● ●).

You have to decide how many of the 5 tokens you place in the **blue bucket** and place in the **yellow bucket**.

Tokens placed in the **blue bucket** are worth **0.5 points**.

Tokens place in the **yellow bucket** are worth **1 point**.

The rule is to place the tokens in the blue bucket.

In summary
Your income in **points** in this task =
the number of tokens you place in the **yellow bucket**
plus
half the number of tokens you place in the **blue bucket**

Public Good Game Instructions

Your task

At the start of this task each participant receives an **endowment** of 5 tokens (●●●●●).

You will be in a group of 4 all of whom are Prolific participants as you are.

These instructions are the same for all group members.

All group members have to decide how many of the 5 tokens they place in the **blue bucket** and place in the **yellow bucket**.

Tokens placed in the **blue bucket** by any group member are **contributed to the group project**.

Tokens place in the **yellow bucket** by any group member are **kept for themselves**.

The other 3 members of your group also make this decision during the experiment.

After all 4 group members have made their individual decision, all tokens placed in the **blue bucket** by any group member are added up, and this number is **doubled** to give the value of the group project.

Each group member gets a **quarter** of the value of the group project, plus the value of what is kept for themselves (1 point for each token they place in the **yellow bucket**).

The rule is to place the tokens in the blue bucket *(T3 – PGG with Rule).*

The rule is to place the tokens in the blue bucket because it helps the other group members *(T4 – PGG with Social Rule).*

In summary
Each group member's income in **points** in this task =
the number of tokens they place in the **yellow bucket**
plus
their share (that is, one quarter) of double the number of tokens anyone places in the **blue bucket**

B Experiment Screenshots

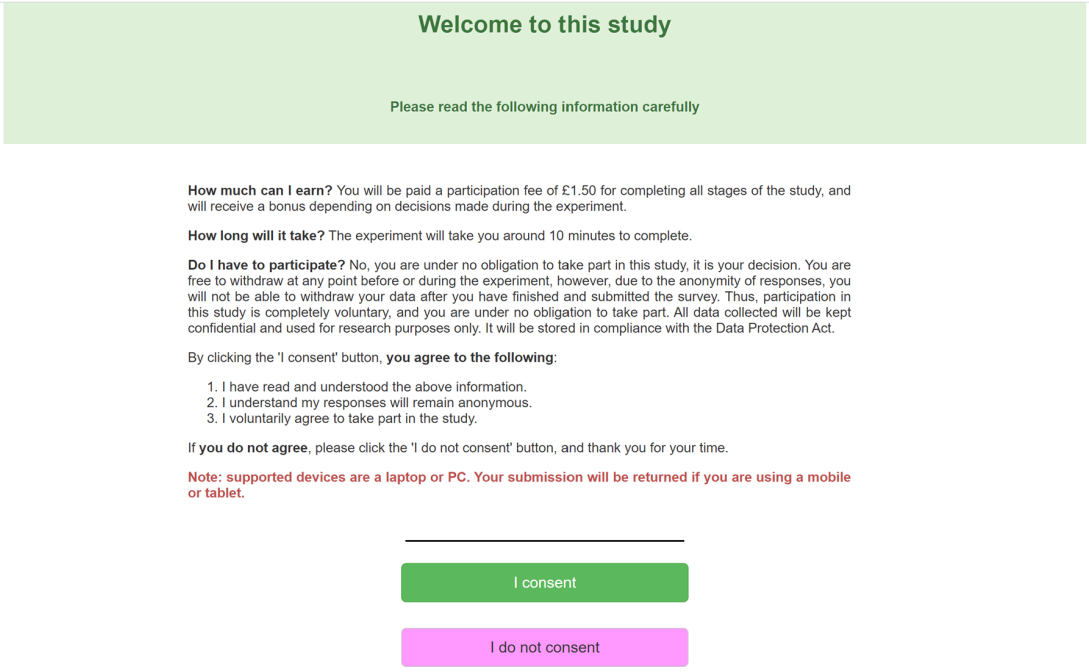

Figure B1: Consent Form

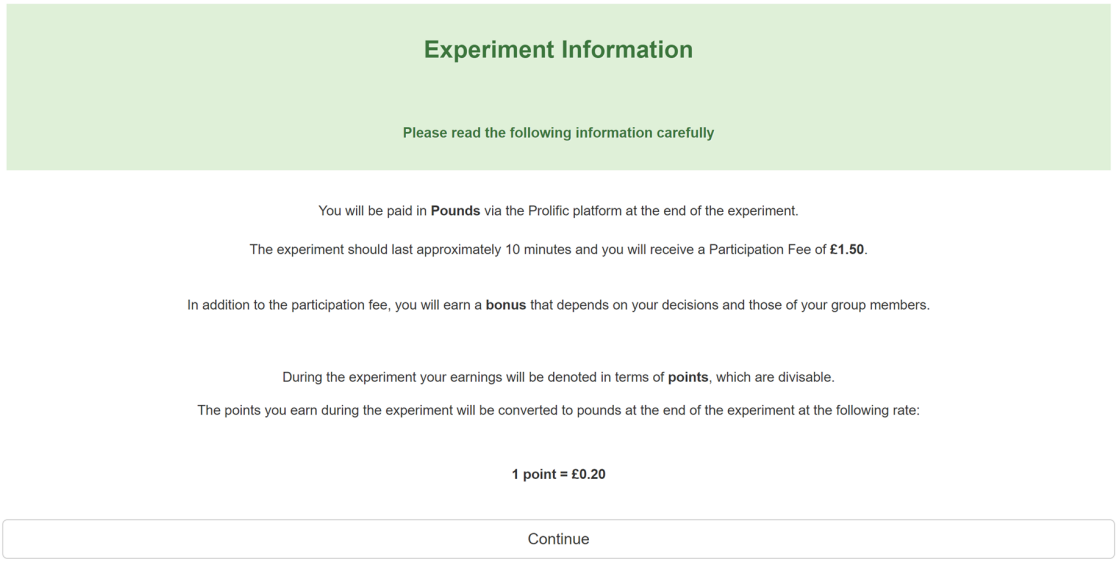

Figure B2: Experiment Information - Baseline PGG, PGG with Rule, and PGG with Social Rule

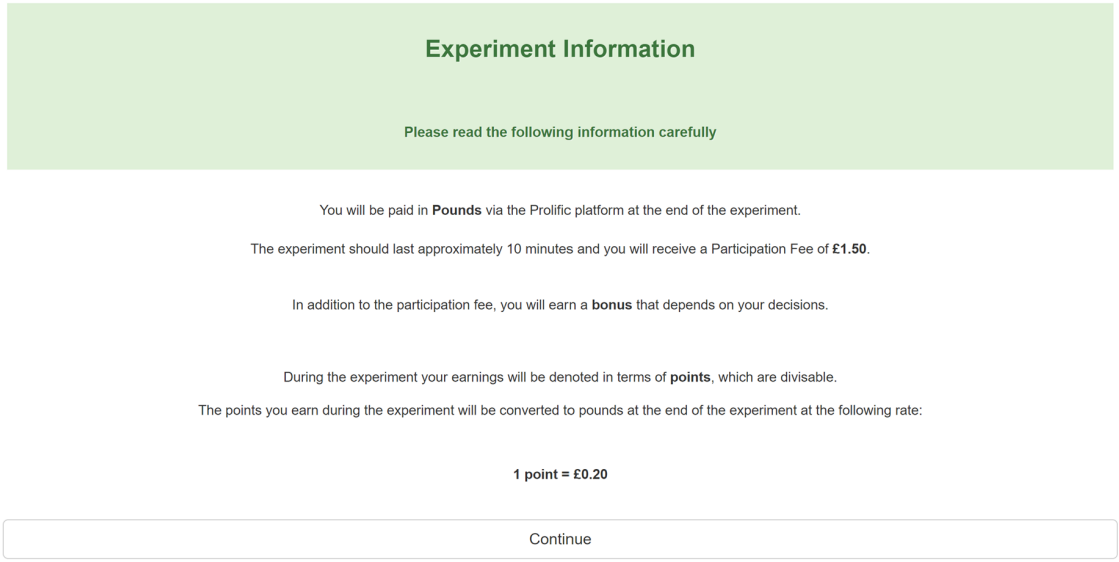

Figure B3: Experiment Information - Baseline RFT

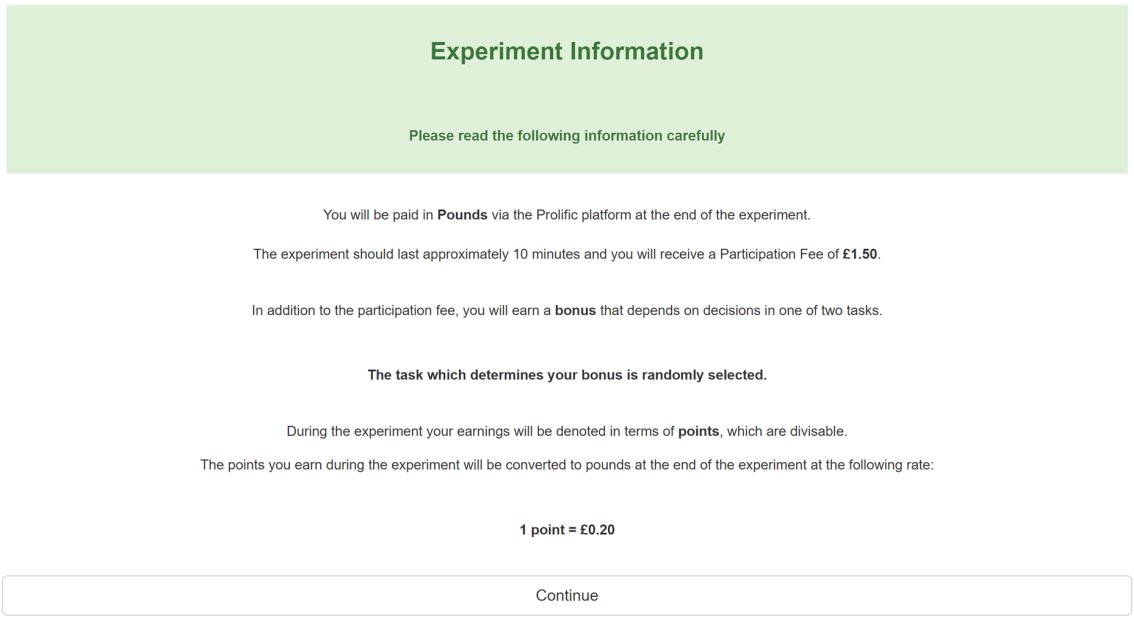

Figure B4: Experiment Information - RFT and PGG

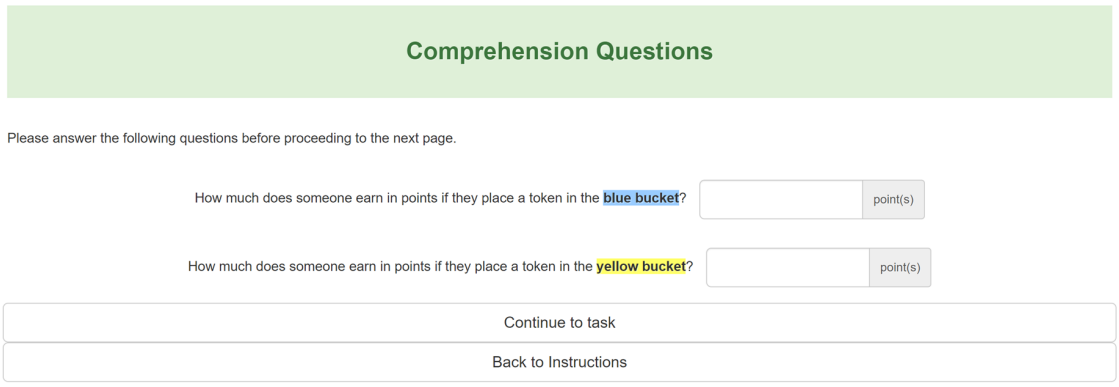

Figure B5: RFT Comprehension Questions

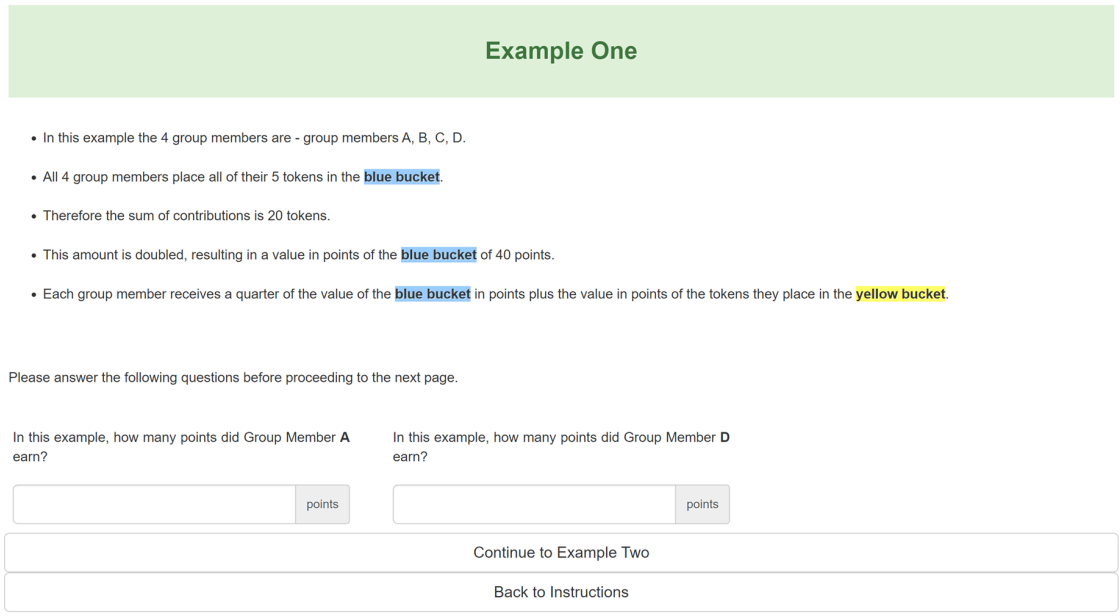

Figure B6: PGG Example One

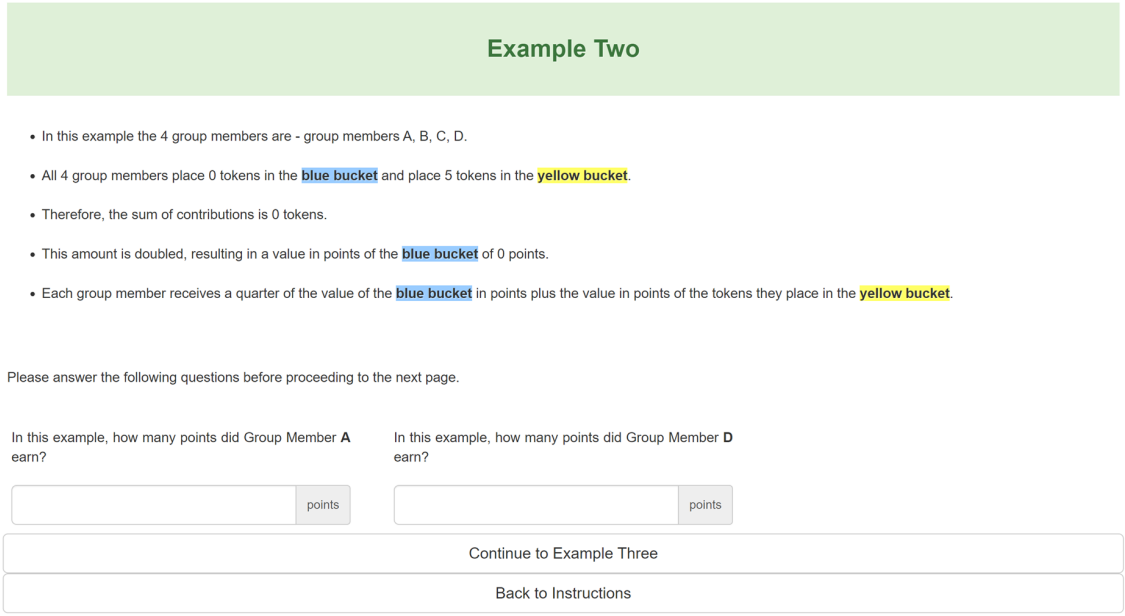

Figure B7: PGG Example Two

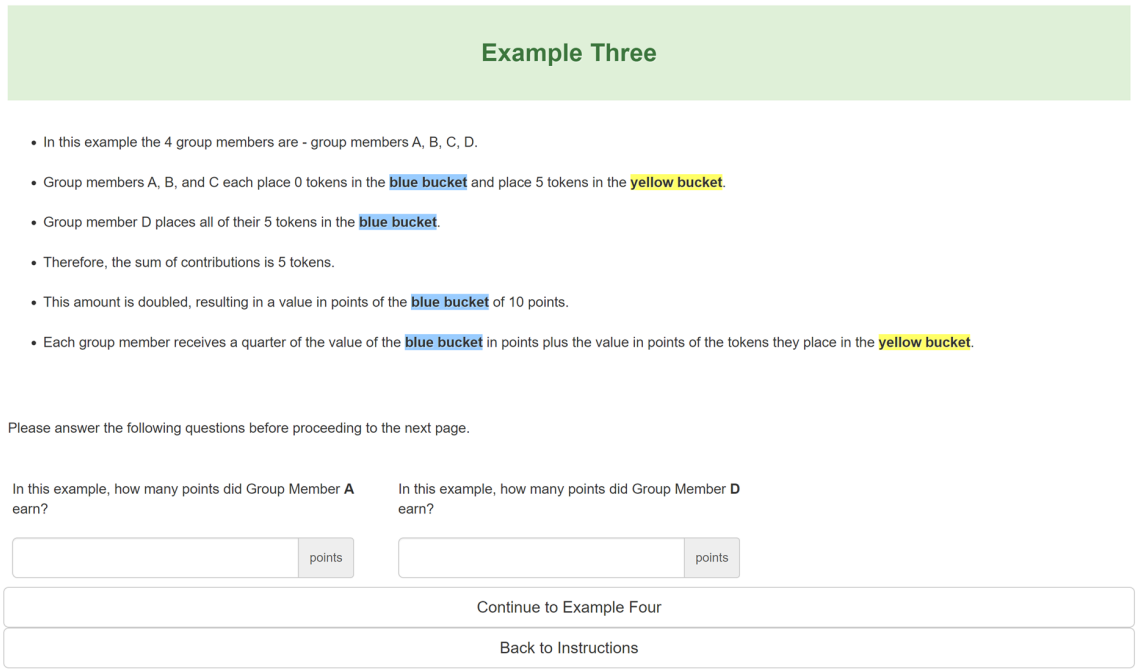

Figure B8: PGG Example Three

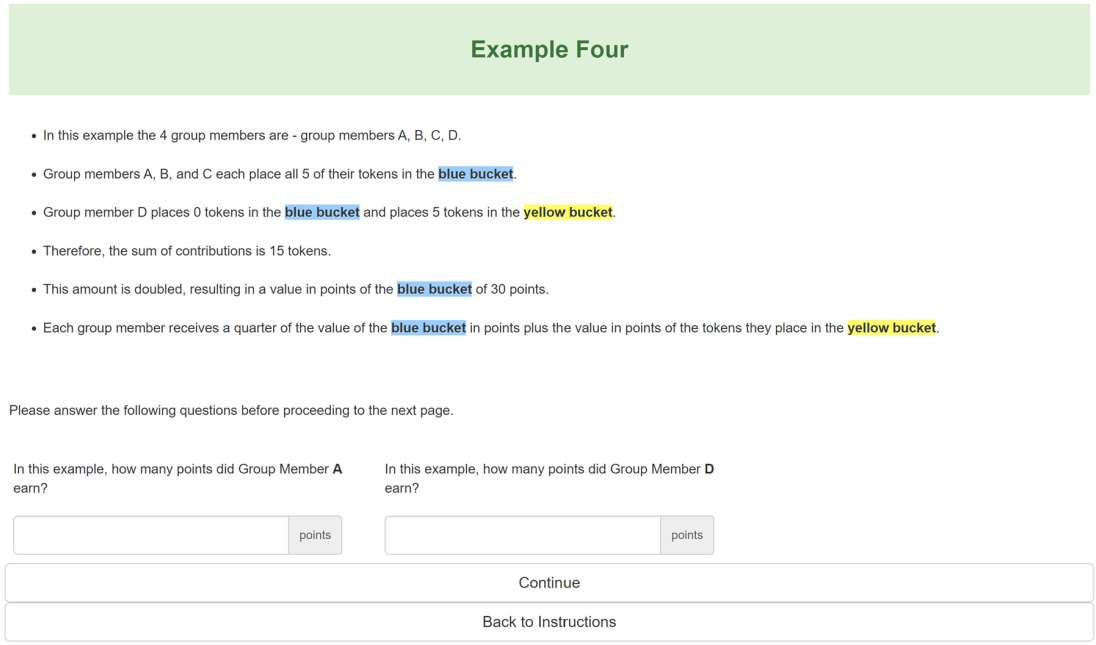

Figure B9: PGG Example Four

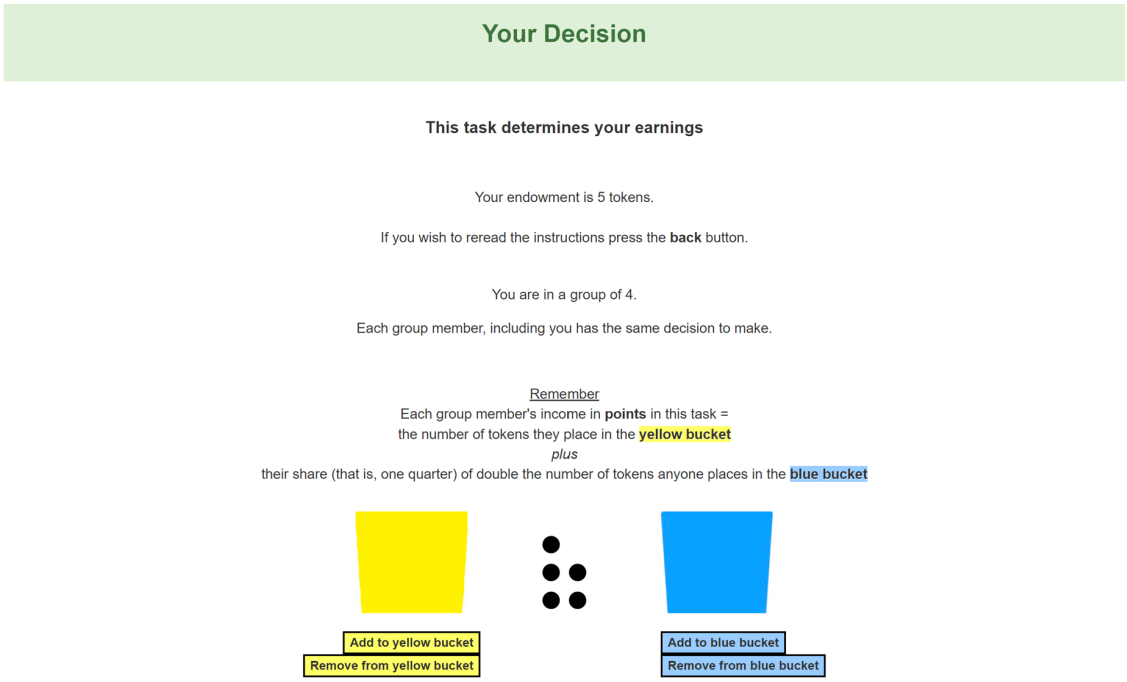

Figure B10: Decision Screen Baseline PGG

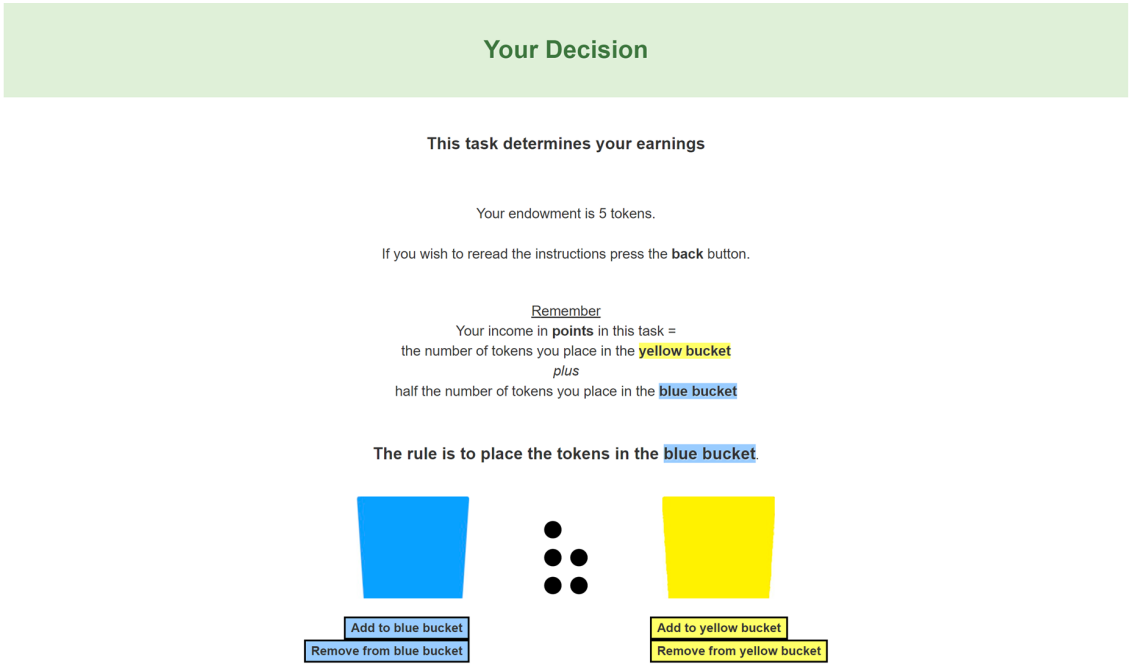

Figure B11: Decision Screen Baseline RFT

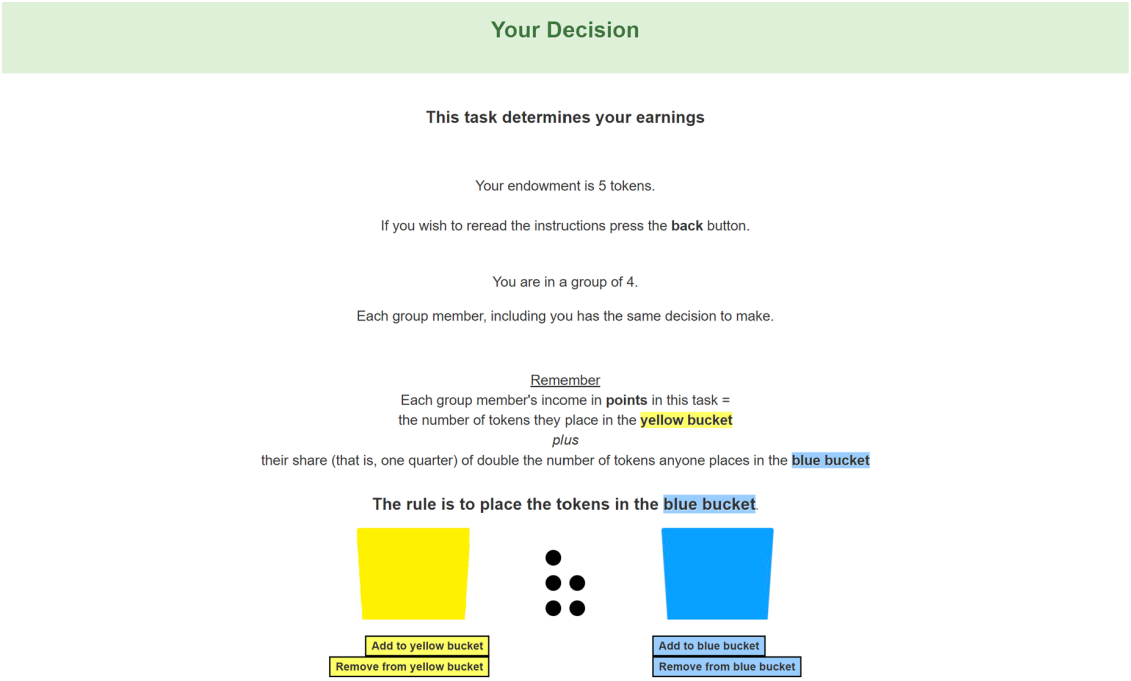

Figure B12: Decision Screen PGG with Rule

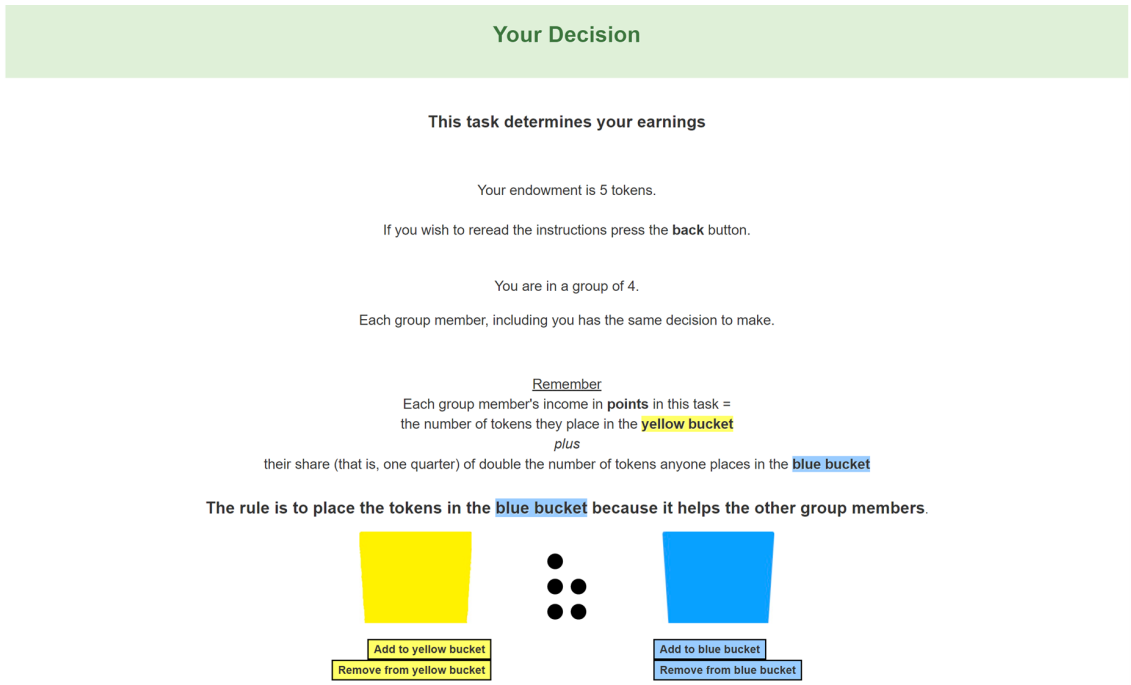

Figure B13: Decision Screen PGG with Social Rule

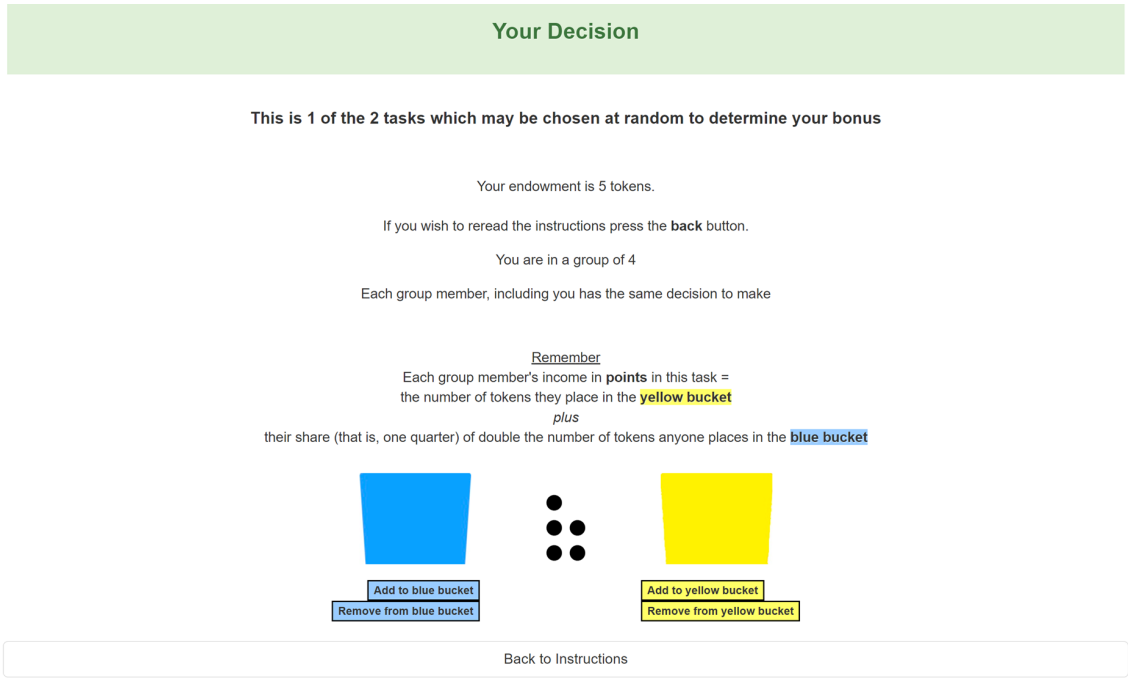

Figure B14: Decision Screen PGG (T5)

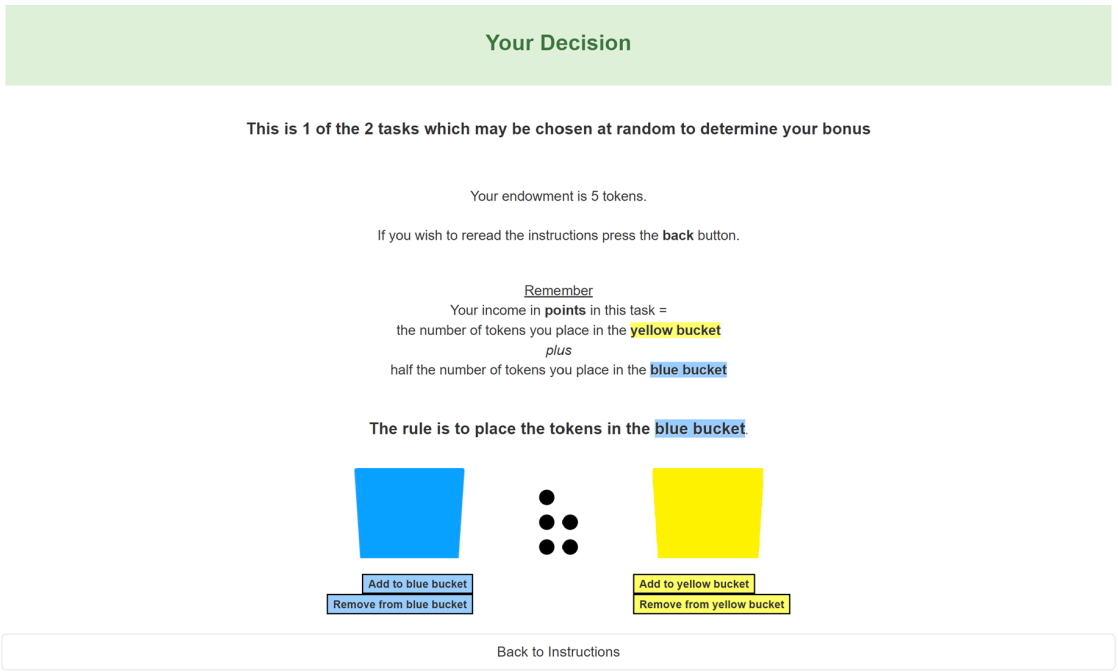

Figure B15: Decision Screen RFT (T5)

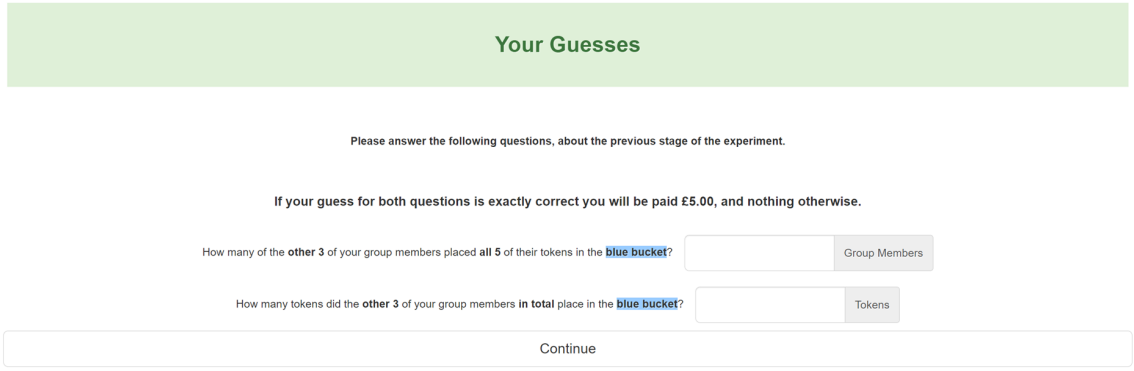

Figure B16: Empirical Expectations Baseline PGG, PGG with Rule, and PGG with Social Rule

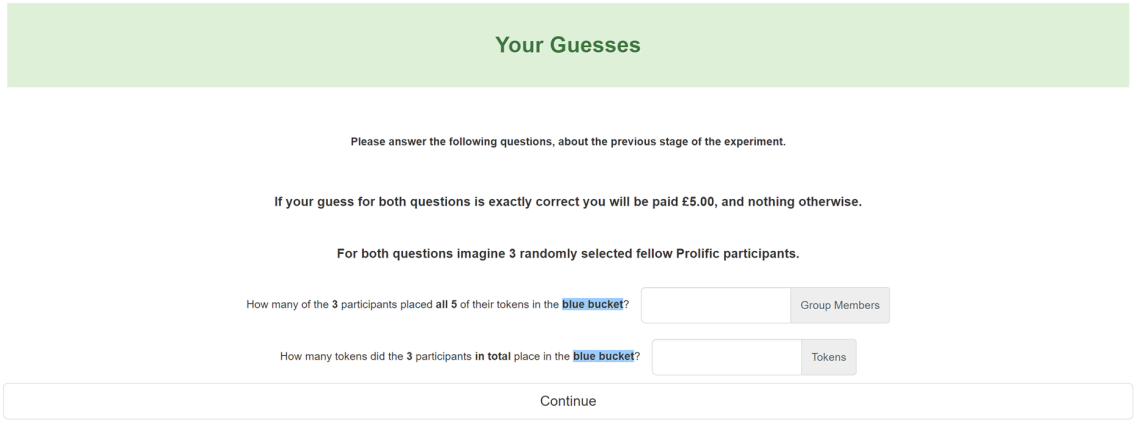

Figure B17: Empirical Expectations Baseline RFT

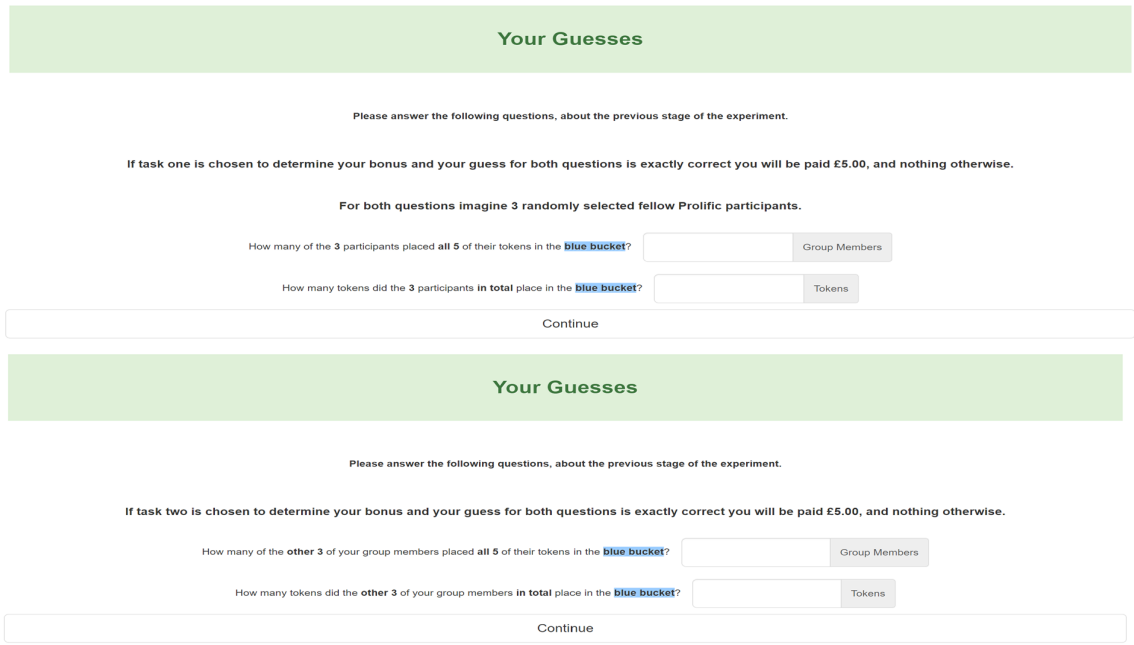

Figure B18: Empirical Expectations RFT and PGG

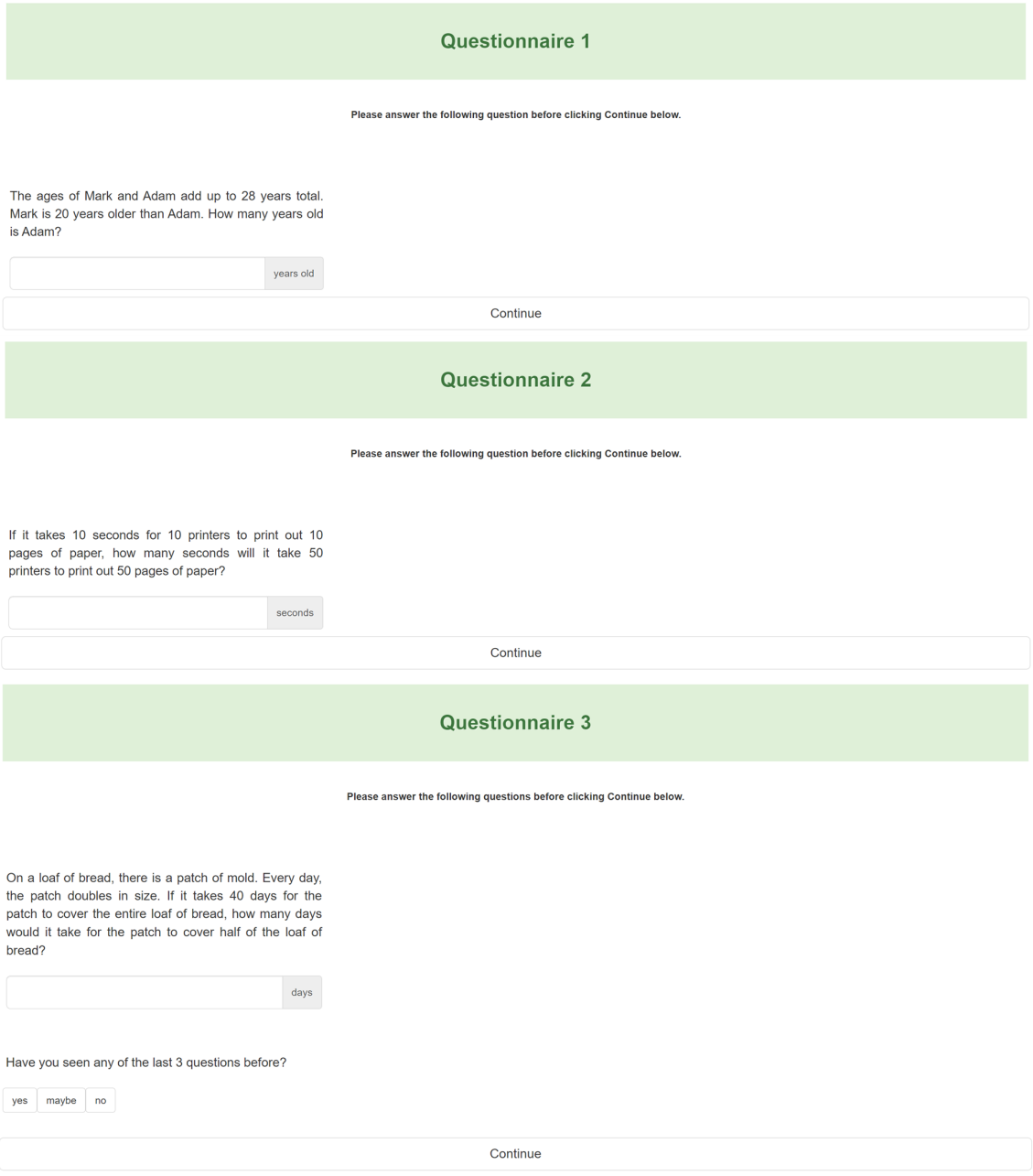

Figure B19: Cognitive Reflection Test

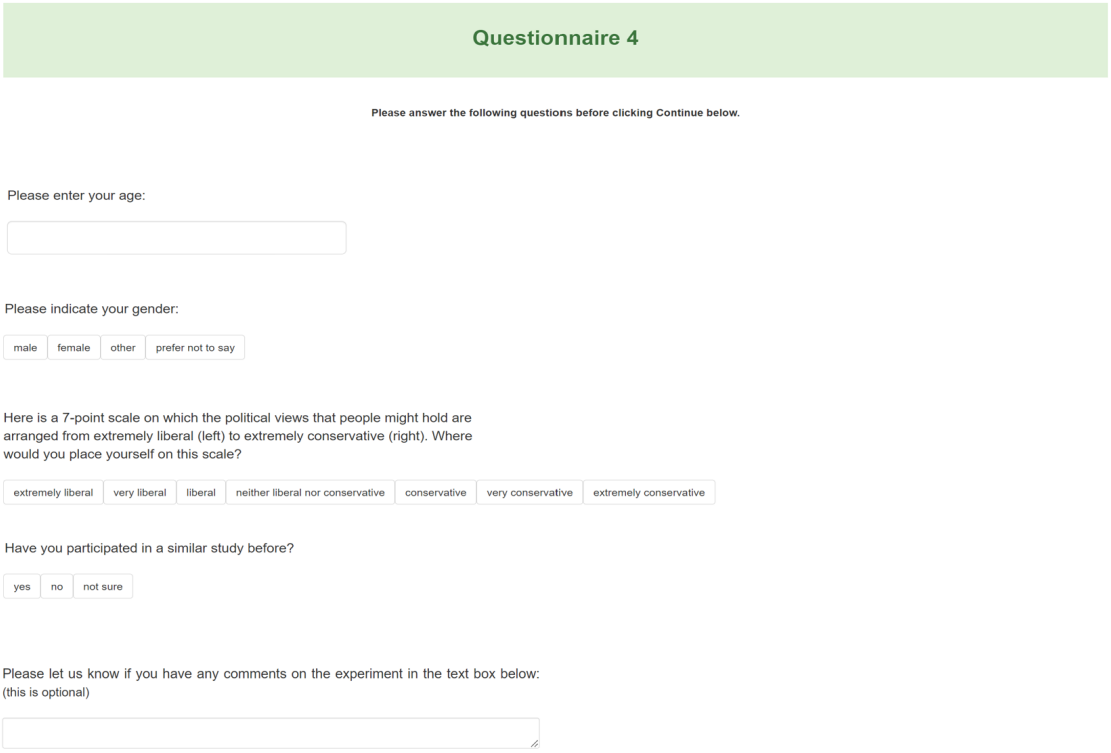

Figure B20: Exit Questionnaire

C Additional Results

As specified in the preregistration analysis, as a robustness check we additionally run a Poisson regression for the regression specifications in the Results and Discussion section. Results reported in Section 5 are unchanged.

Poisson Regressions – Reporting Incident Rate Ratios

Table C1: Rule following behaviour

	Rate of rule following		
	(1)	*(2)*	*(3)*
T3	1.071	1.074	1.035
	(0.0688)	(0.0686)	(0.0561)
T4	1.142*	1.144*	1.057
	(0.0678)	(0.0666)	(0.0512)
CRT	0.994	1.000	0.988
	(0.0217)	(0.0232)	(0.0193)
Constant	3.505***	3.325***	2.396***
	(0.196)	(0.455)	(0.306)
N	309	308	308
Additional covariates	None	Demographics	Demographics + Empirical Expectations

All results are from Poisson regressions of the rate of rule following (as measured by the number of tokens placed in the blue bucket), with T2 as the omitted baseline, in the first column on a dummy variable for T3, a dummy variable for T4, and CRT score. Regressions in column 2 include demographic controls. Regressions in column 3 include demographic controls and empirical expectations. Coefficients represent incident rate ratios. Robust standard errors in parentheses. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05

Table C2: Cooperative behaviour

	Contributions to public good		
	(1)	*(2)*	*(3)*
T3	1.190**	1.200**	1.125*
	(0.0722)	(0.0730)	(0.0571)
T4	1.270***	1.297***	1.163**
	(0.0703)	(0.0706)	(0.0535)
CRT	0.996	0.997	1.001
	(0.0204)	(0.0219)	(0.0175)
Constant	3.147***	2.720***	2.215***
	(0.158)	(0.359)	(0.283)
N	296	293	293
Additional covariates	None	Demographics	Demographics + Empirical Expectations

All results are from Poisson regressions of the rate of contributions to the public good (as measured by the number of tokens placed in the blue bucket), with T1 as the omitted baseline, in the first column on a dummy variable for T3, a dummy variable for T4, and CRT score. Regressions in column 2 include demographic controls. Regressions in column 3 include demographic controls and empirical expectations. Coefficients represent incident rate ratios. Robust standard errors in parentheses. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05

Table C3: Relationship between rule following and cooperation

	Tokens placed in the blue bucket		
	(1)	*(2)*	*(3)*
PGG Dummy	0.790**	0.792**	0.855*
	(0.0629)	(0.0630)	(0.0628)
Order Dummy	0.663**	0.670**	0.819*
	(0.0992)	(0.0973)	(0.0758)
PGG × Order	1.575**	1.569**	1.223
	(0.262)	(0.260)	(0.141)
CRT	0.986	0.966	1.004
	(0.0427)	(0.0404)	(0.0287)
Constant	3.758***	3.663***	1.751*
	(0.400)	(0.859)	(0.394)
N	222	222	220
Additional covariates	None	Demographics	Demographics + Empirical Expectations

Robust standard errors clustered at the level of the individual. All results are from Poisson regressions of the number of tokens placed in the blue bucket in T5, in the first column on a dummy variable for whether it was a PGG, a dummy variable for the order in which tasks were completed, an interaction effect between the PGG dummy and the order dummy, and CRT score. Regressions in column 2 include demographic controls. Regressions in column 3 include demographic controls and empirical expectations. Coefficients represent incident rate ratios. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05

Table C4: The role of empirical expectations

	Tokens placed in the blue bucket							
	(T1 1)	*(T1 2)*	*(T2 1)*	*(T2 2)*	*(T3 1)*	*(T3 2)*	*(T4 1)*	*(T4 2)*
EmpExpAll	1.229***	1.216***	1.212***	1.179**	1.166***	1.170***	1.130***	1.148***
	(0.0439)	(0.0387)	(0.0664)	(0.0695)	(0.0449)	(0.0462)	(0.0340)	(0.0333)
EmpExpTotal	1.010	1.014	1.037***	1.042***	1.015	1.019*	1.008	1.009
	(0.00843)	(0.00741)	(0.0108)	(0.0110)	(0.00871)	(0.00869)	(0.00650)	(0.00626)
CRT	0.980	0.985	0.941	0.937	1.020	1.030	1.009	0.998
	(0.0285)	(0.0275)	(0.0346)	(0.0363)	(0.0357)	(0.0351)	(0.0238)	(0.0275)
Constant	2.074***	1.563	1.922***	1.934**	2.305***	2.663***	2.783***	2.704***
	(0.207)	(0.401)	(0.262)	(0.465)	(0.281)	(0.567)	(0.263)	(0.551)
N	104	102	117	117	93	92	99	99
Additional covariates	None	Demographics	None	Demographics	None	Demographics	None	Demographics

Results from Poisson regressions of the number of tokens placed in the blue bucket on empirical expectations (i.e., number of participants who placed all 5 of their tokens in the blue bucket (EmpExpAll) and total number of tokens placed in the blue bucket by other participants (EmpExpTotal), and CRT score. Column 2 for each treatment includes demographic controls. Coefficients represent incident rate ratios. Robust standard errors in parentheses. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05

Table C5: The role of empirical expectations in T5

	Tokens placed in the blue bucket			
	(RFT 1)	*(RFT 2)*	*(PGG 1)*	*(PGG 2)*
EmpExpAll	1.354***	1.356***	1.292***	1.318***
	(0.0959)	(0.0984)	(0.0657)	(0.0643)
EmpExpTotal	1.048***	1.049***	1.034**	1.031*
	(0.0143)	(0.0147)	(0.0117)	(0.0127)
Order	0.859	0.820*	1.027	1.007
	(0.0772)	(0.0784)	(0.0863)	(0.0840)
CRTScore	1.023	0.996	1.023	1.005
	(0.0361)	(0.0362)	(0.0412)	(0.0393)
Constant	1.150	1.546	1.266	1.821*
	(0.173)	(0.476)	(0.216)	(0.437)
N	111	111	109	109
Additional covariates	None	Demographics	None	Demographics

Results from Poisson regressions of the number of tokens placed in the blue bucket on empirical expectations (i.e., number of participants who placed all 5 of their tokens in the blue bucket (EmpExpAll) and total number of tokens placed in the blue bucket by other participants (EmpExpTotal), and CRT score. Columns 1 and 2 refer to the RFT with column 2 including demographic controls. Columns 3 and 4 refer to the PGG with column 4 including demographic controls. Coefficients represent incident rate ratios. Robust standard errors in parentheses. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05

Table C6: Empirical Expectations All

	The role of empirical expectations	
	(EmpExpAll 1)	*(EmpExpAll 2)*
T2	1.101	1.162
	(0.106)	(0.111)
T3	1.179	1.193
	(0.120)	(0.123)
T4	1.345**	1.383***
	(0.122)	(0.125)
CRT	0.970	0.980
	(0.0278)	(0.0297)
Constant	1.623***	0.987
	(0.141)	(0.217)
N	413	410
Additional covariates	None	Demographics

Results from Poisson regressions of empirical expectations (number of participants who placed all of their tokens in the blue bucket (EmpExpAll) and total number of tokens placed in blue bucket by other participants (EmpExpTotal), on Treatment dummies and CRT score. T1 serves as the omitted baseline. Coefficients represent incident rate ratios. Robust standard errors in parentheses. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05

Table C7: Empirical Expectations Total

	The role of empirical expectations	
	(EmpExpTotal 1)	*(EmpExpTotal 2)*
T2	1.000	1.018
	(0.0844)	(0.0869)
T3	1.187*	1.170
	(0.0955)	(0.0974)
T4	1.236**	1.225**
	(0.0943)	(0.0959)
CRT	1.053*	1.045
	(0.0270)	(0.0279)
Constant	7.560***	7.636***
	(0.572)	(1.380)
N	413	410
Additional covariates	None	Demographics

Results from OLS regressions of empirical expectations (number of participants who placed all of their tokens in the blue bucket (EmpExpAll) and total number of tokens placed in blue bucket by other participants (EmpExpTotal), on Treatment dummies and CRT score. T1 serves as the omitted baseline. Coefficients represent incident rate ratios. Robust standard errors in parentheses. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05

Table C8: Empirical Expectations All T5

	The role of empirical expectations	
	(EmpExpAll 1)	*(EmpExpAll 2)*
PGG Dummy	0.846	0.848
	(0.0832)	(0.0830)
Order Dummy	0.702*	0.737
	(0.117)	(0.120)
PGG × Order	1.553*	1.558*
	(0.298)	(0.297)
CRT	0.902	0.899*
	(0.0480)	(0.0462)
Constant	2.167***	1.108
	(0.258)	(0.405)
N	220	220
Additional covariates	Demographics	Demographics

Robust standard errors clustered at the level of the individual in parentheses. All results are from ordinary least squares regressions of empirical expectations (i.e., number of participants who placed all 5 of their tokens in the blue bucket (EmpExpAll) and total number of tokens placed in the blue bucket by other participants (EmpExpTotal),on a dummy variable for whether it was a PGG, a dummy variable for the order in which tasks were completed, an interaction effect between the PGG dummy and the order dummy, and CRT score. Regressions include demographic controls. Coefficients represent incident rate ratios. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05

Table C9: Empirical Expectations Total T5

	The role of empirical expectations	
	(EmpExpTotal 1)	*(EmpExpTotal 2)*
PGG Dummy	0.936	0.939
	(0.0731)	(0.0730)
Order Dummy	0.643**	0.670*
	(0.107)	(0.107)
PGG × Order	1.496*	1.493*
	(0.267)	(0.267)
CRT	1.015	1.003
	(0.0506)	(0.0449)
Constant	9.183***	7.461***
	(1.168)	(1.694)
N	222	222
Additional covariates	Demographics	Demographics

Robust standard errors clustered at the level of the individual in parentheses. All results are from ordinary least squares regressions of empirical expectations (i.e., number of participants who placed all 5 of their tokens in the blue bucket (EmpExpAll) and total number of tokens placed in the blue bucket by other participants (EmpExpTotal),on a dummy variable for whether it was a PGG, a dummy variable for the order in which tasks were completed, an interaction effect between the PGG dummy and the order dummy, and CRT score. Regressions include demographic controls. Coefficients represent incident rate ratios. Significance levels are denoted by: *** p<.001, ** p<.01, * p<.05

Table C10: Demographic Characteristics ($N = 525$)

		Baseline PGG	Baseline RFT	PGG with Rule	PGG with Social Rule	RFT and PGG	
						Order 0	Order 1
Average Age (SD)		38.86 (12.71)	36.91 (12.29)	39.73 (13.55)	38.45 (12.92)	36.29 (11.65)	37.49 (13.96)
Gender	Male	0.33	0.34	0.39	0.38	0.37	0.25
	Female	0.67	0.65	0.60	0.61	0.61	0.75
	Other	–	–	0.01	–	0.02	–
	Prefer Not to Say	–	0.01	–	0.01	–	–
Political Ideology	Liberal	0.56	0.50	0.58	0.47	0.57	0.51
	Conservative	0.12	0.09	0.20	0.21	0.17	0.15
	Neither	0.32	0.41	0.23	0.31	0.26	0.34
Similar Study	Yes	0.15	0.11	0.10	0.11	0.09	0.15
	No	0.70	0.75	0.75	0.77	0.75	0.72
	Not sure	0.15	0.14	0.15	0.12	0.15	0.13

† RFT – Rule Following Task, PGG – Public Goods Game, SD – Standard Deviation

‡ Characteristics are for the final sample of 525 participants

* For ease of exposition political ideology is collapsed into three categories: 1. liberal if the participant indicated they were either extremely liberal, very liberal, or liberal; 2. conservative if the participant indicated they were conservative, very conservative, or extremely conservative; and 3. neither if the participant indicated they were neither conservative nor liberal. Gender, Political Ideology, and Similar Study show the percentage of participants in each category. Due to rounding totals may not sum to 1.

Graphs

The additional graphs below present more granular breakdowns of those reported in the main text. Numbers below the green bars refer to percentage of participants in that category.

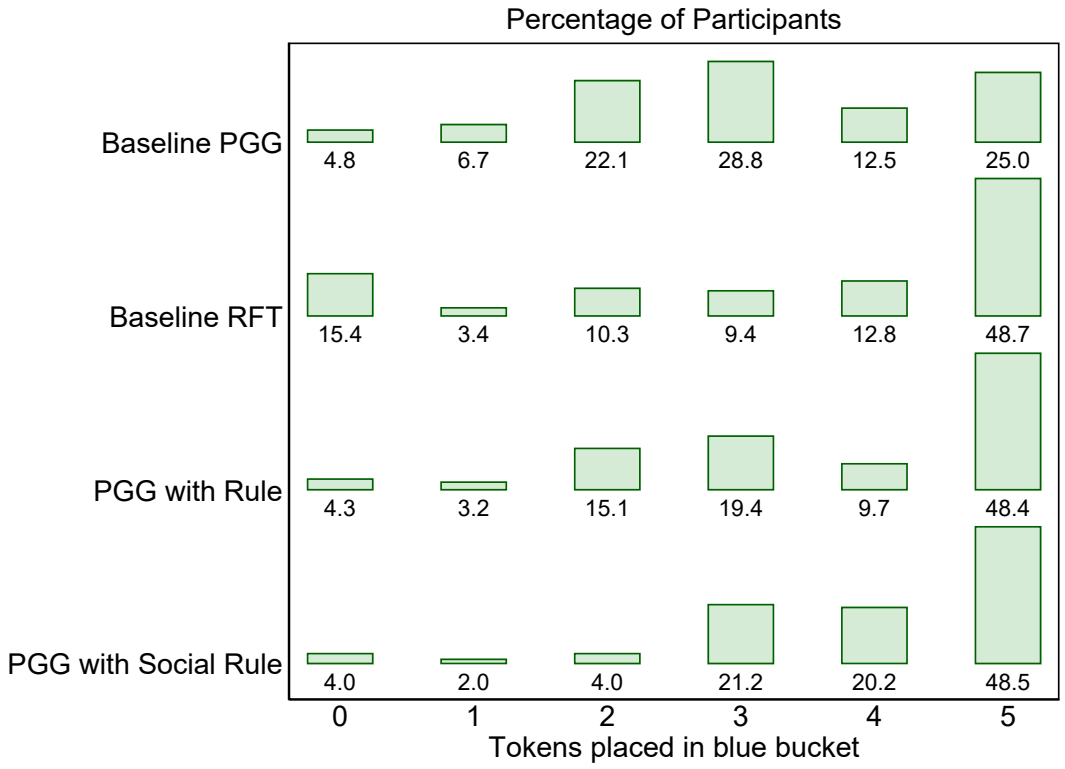

Figure C1: Between subjects treatments – distribution of tokens placed in the blue bucket

The green bars show the percentage of participants in each treatment who placed a given number of tokens (as indicated by the x axis) in the blue bucket. The numbers below each bar denote the corresponding percentage rounded to one decimal place.

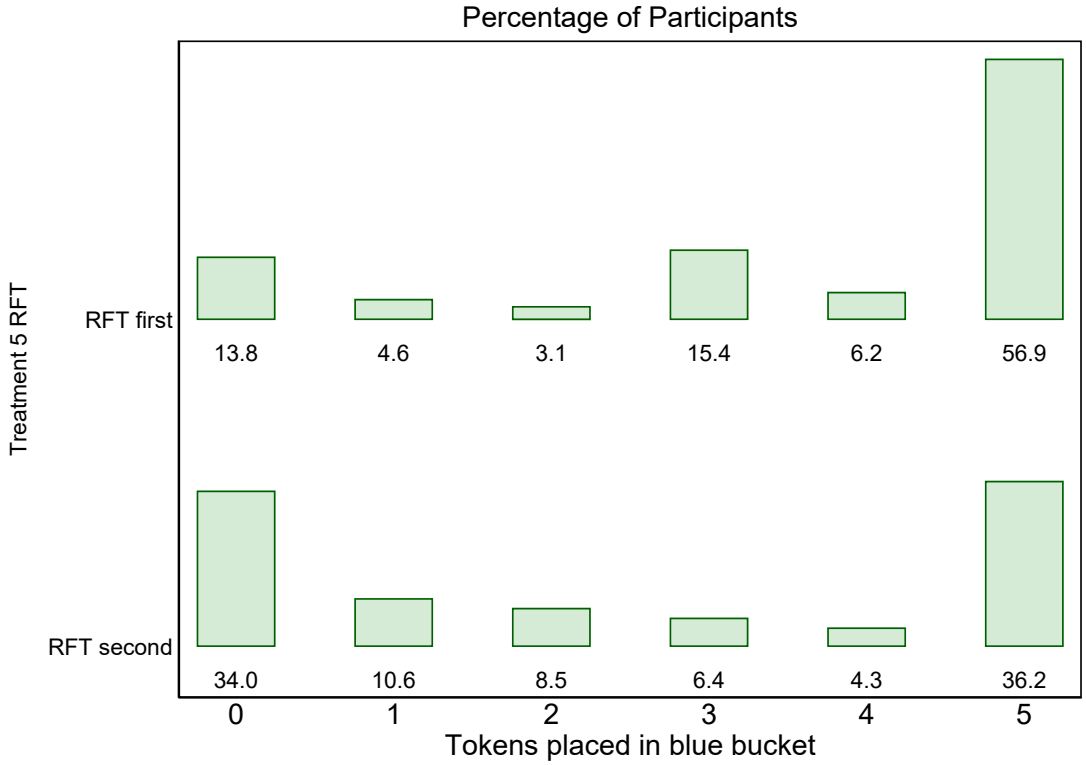

Figure C2: T5 RFT – distribution of tokens placed in the blue bucket

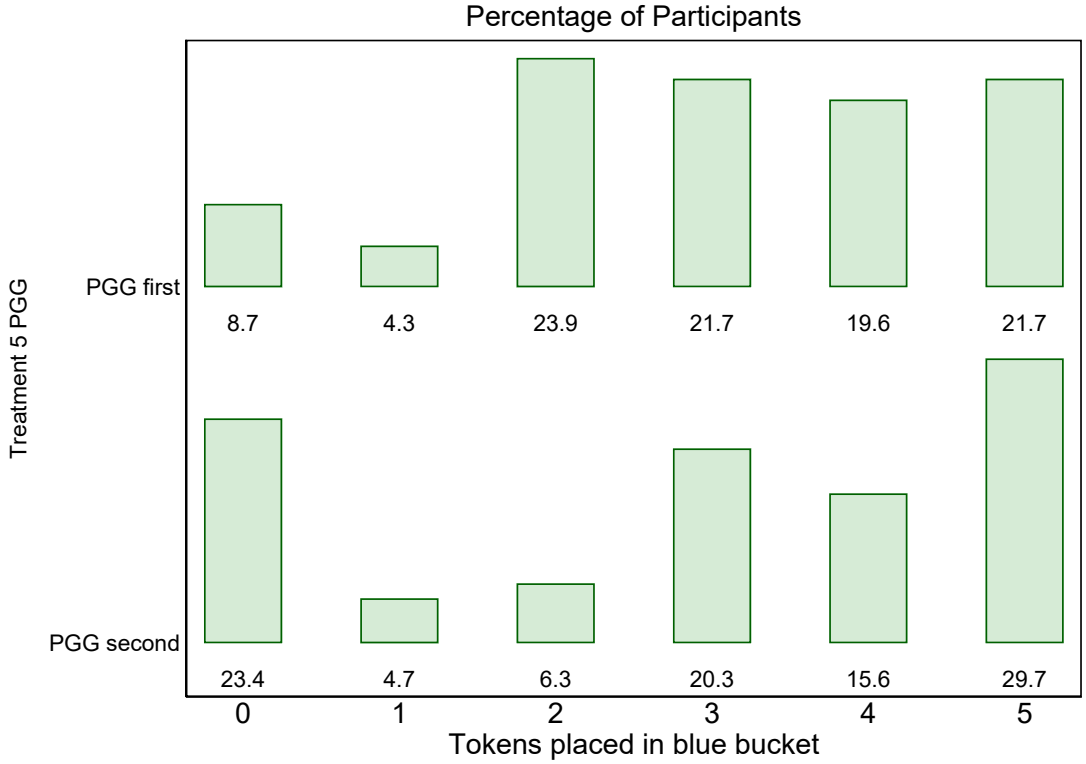

Figure C3: T5 PGG – distribution of tokens placed in the blue bucket

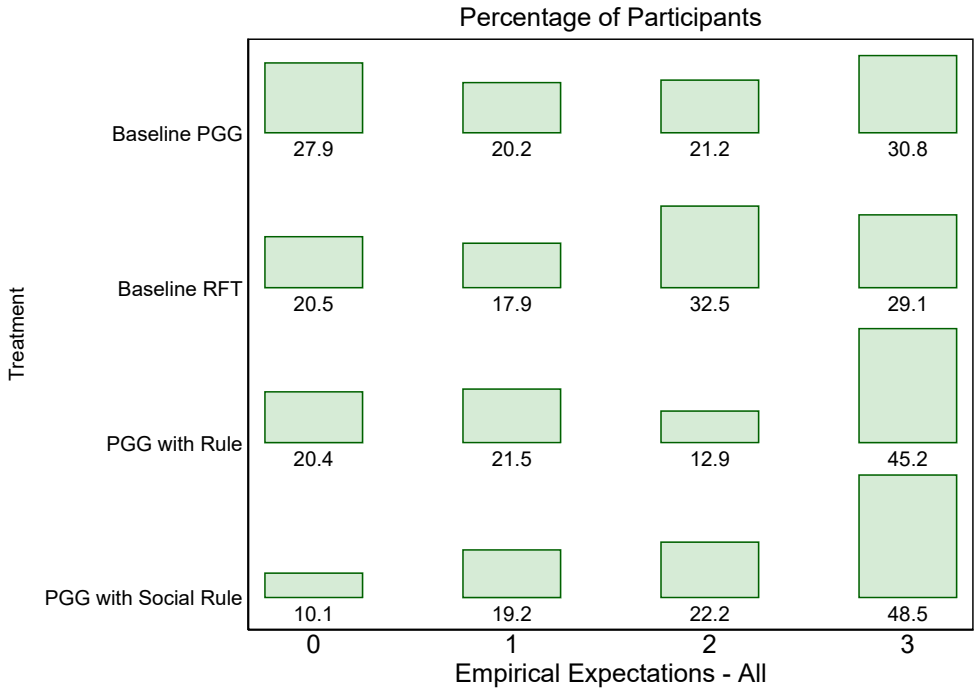

Figure C4: Between subjects treatments – distribution of EmpExpAll

Depending on treatment participants were asked either: *How many of the **other 3** of your group members placed **all 5** of their tokens in the blue bucket?* or to imagine 3 randomly selected fellow Prolific participants and *How many of the **3** participants placed **all 5** of their tokens in the blue bucket?*

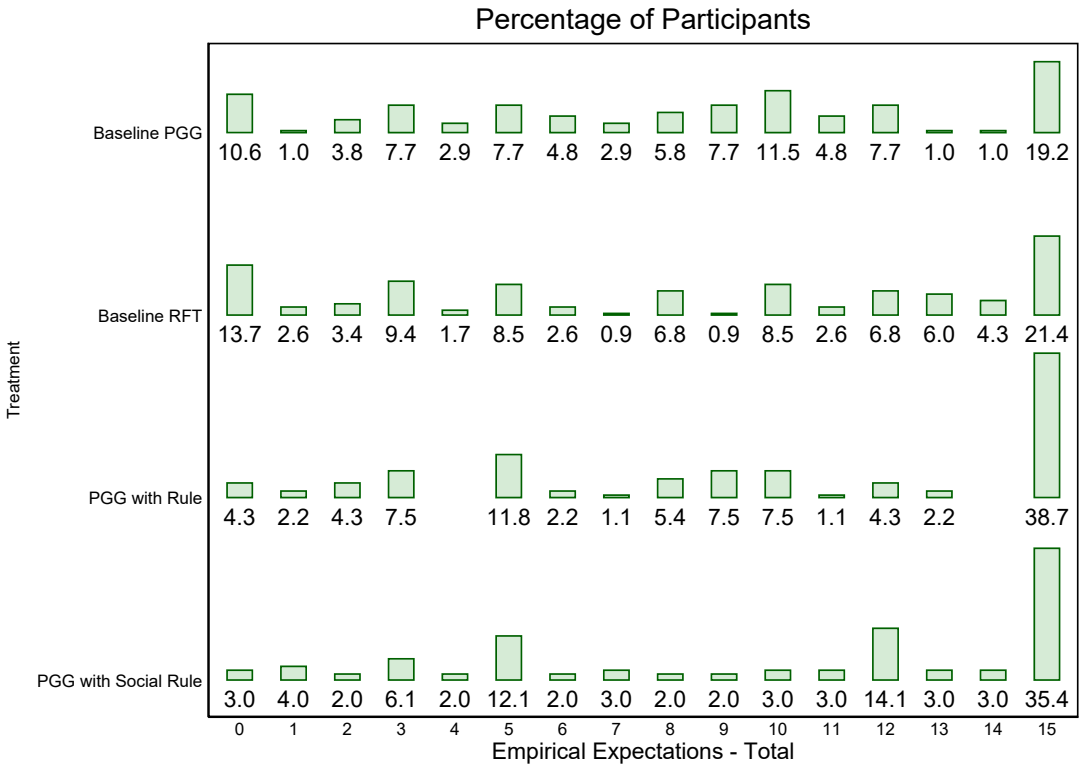

Figure C5: Between subjects treatments – distribution of EmpExpTotal

Depending on treatment participants were asked either: *How many tokens did the **other 3** of your group members **in total** place in the blue bucket?* or to imagine 3 randomly selected fellow Prolific participants and *How many tokens did the **3** participants **in total** place in the blue bucket?*

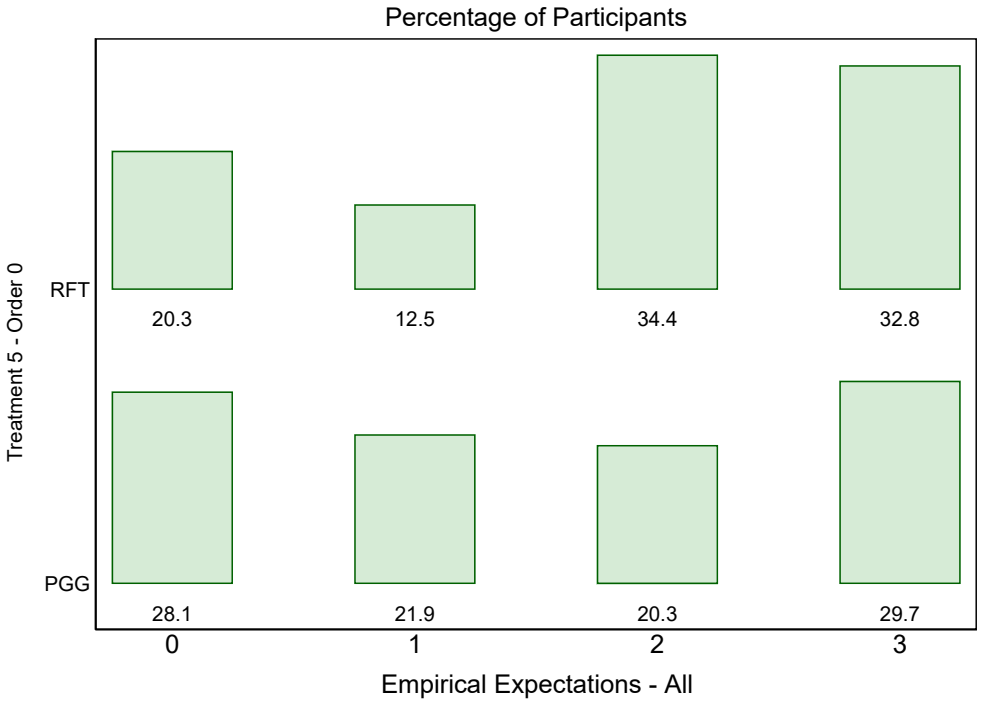

Figure C6: T5 Order 0 – distribution of EmpExpAll

In Order 0 participants completed the RFT first and the PGG second. Depending on the task participants were asked either: *How many of the **other 3** of your group members placed **all 5** of their tokens in the blue bucket?* or to imagine 3 randomly selected fellow Prolific participants and *How many of the **3** participants placed **all 5** of their tokens in the blue bucket?*

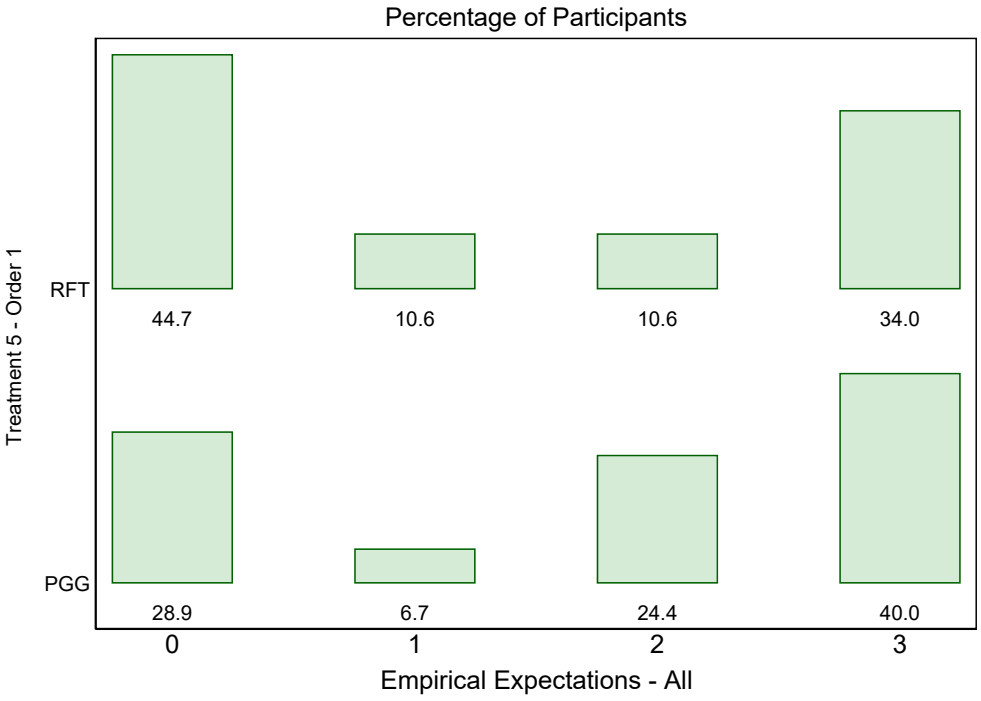

Figure C7: T5 Order 1 – distribution of EmpExpAll

In Order 1 participants completed the PGG first and the RFT second. Depending on the task participants were asked either: *How many of the **other 3** of your group members placed **all 5** of their tokens in the blue bucket?* or to imagine 3 randomly selected fellow Prolific participants and *How many of the **3** participants placed **all 5** of their tokens in the blue bucket?*

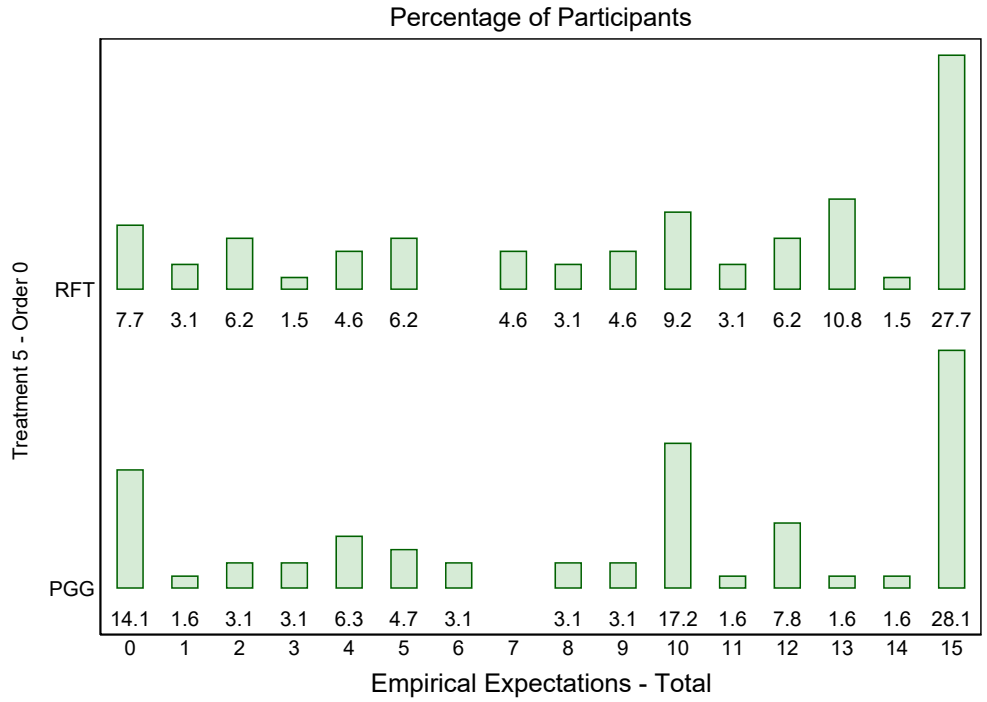

Figure C8: T5 Order 0 – distribution of EmpExpTotal

In Order 0 participants completed the RFT first and the PGG second. Depending on the task participants were asked either: *How many tokens did the **other 3** of your group members **in total** place in the blue bucket?* or to imagine 3 randomly selected fellow Prolific participants and *How many tokens did the **3** participants **in total** place in the blue bucket?*

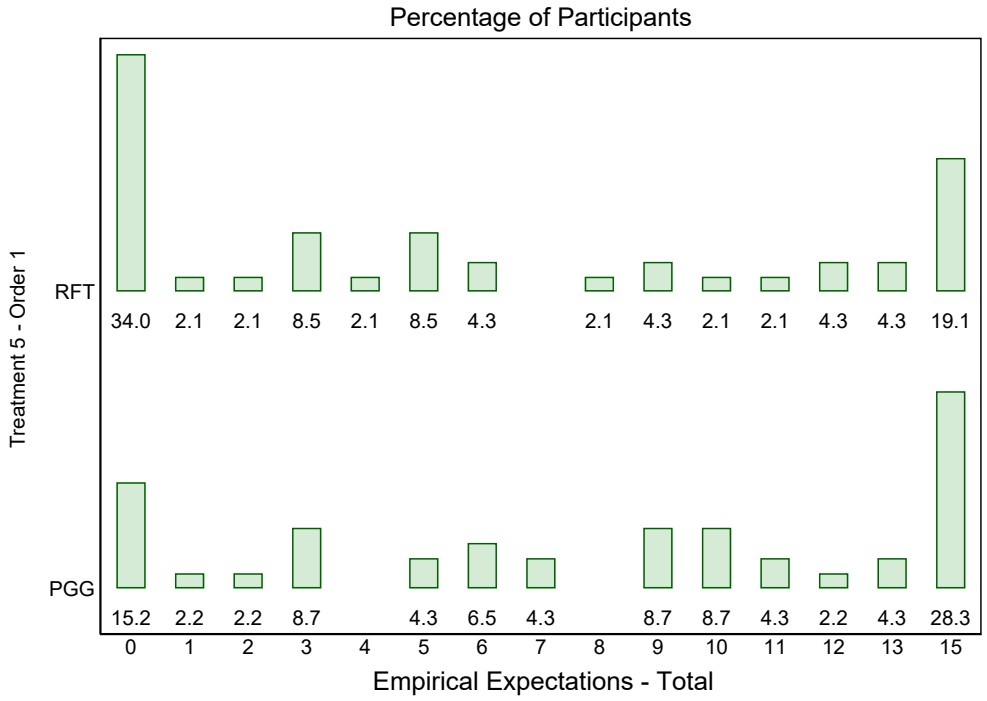

Figure C9: T5 Order 1 – distribution of EmpExpTotal

In Order 1 participants completed the PGG first and the RFT second. Depending on the task participants were asked either: *How many tokens did the **other 3** of your group members **in total** place in the blue bucket?* or to imagine 3 randomly selected fellow Prolific participants and *How many tokens did the **3** participants **in total** place in the blue bucket?*

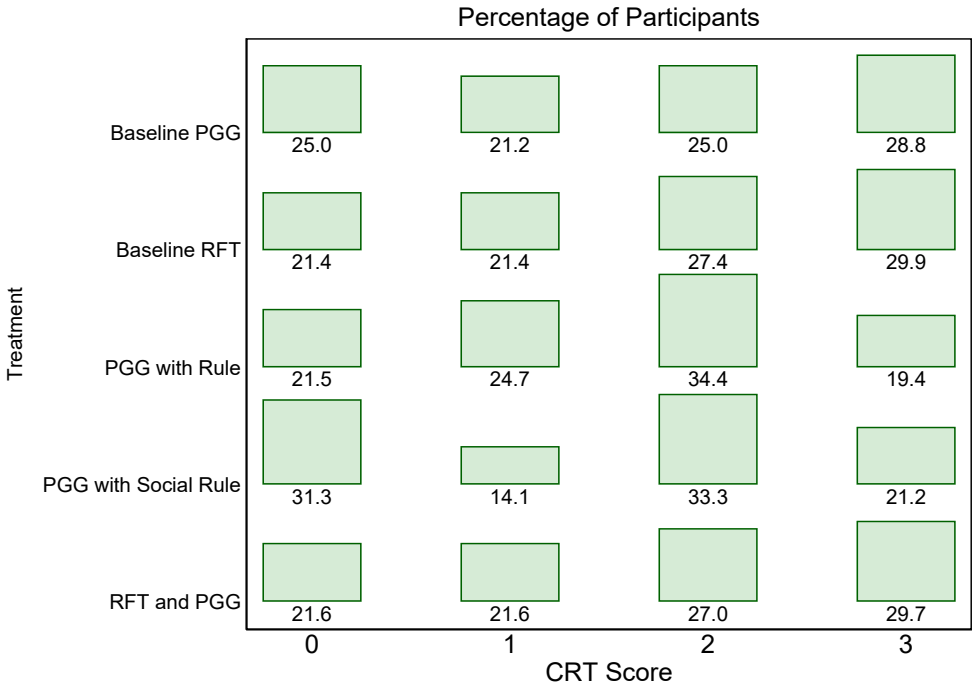

Figure C10: Distribution of CRT scores across the 5 treatments

The green bars show the percentage of participants in each treatment with a given score on the three question Cognitive Reflection Test (CRT). A score of 0 means the participant got all 3 questions wrong, while a score of 3 means the participant got all 3 questions correct. The numbers below each bar denote the corresponding percentage of participants in each category separated by treatment, rounded to one decimal place.

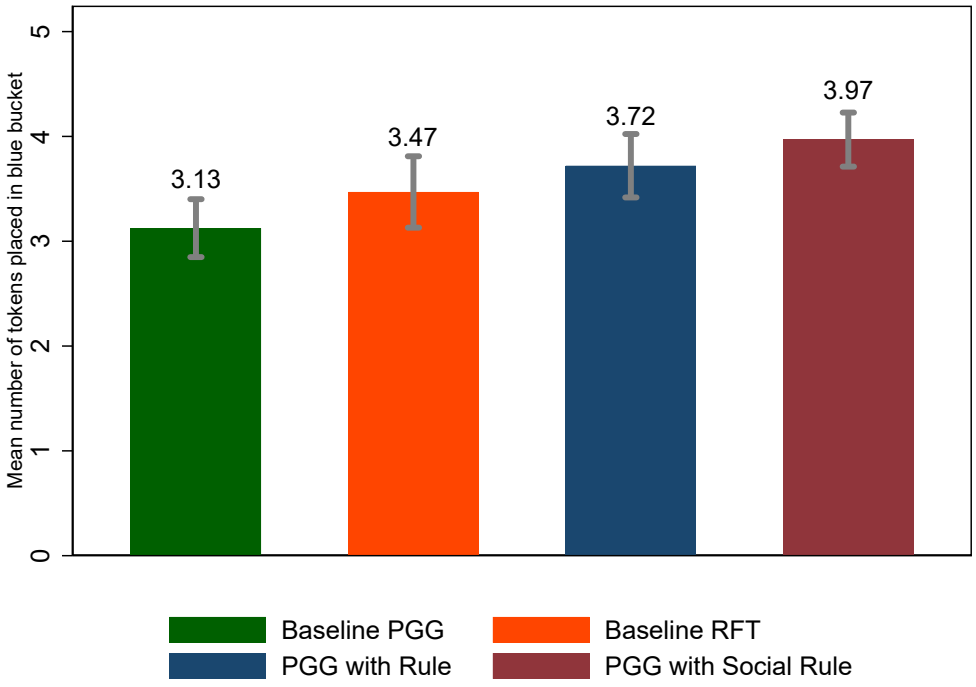

Figure C11: Mean number of tokens placed in the blue bucket for treatments 1 to 4

The above figure shows the mean number of tokens placed in the blue bucket across treatments 1 to 4. The number above each bar corresponds to this mean. Grey bars represent 95% confidence intervals.

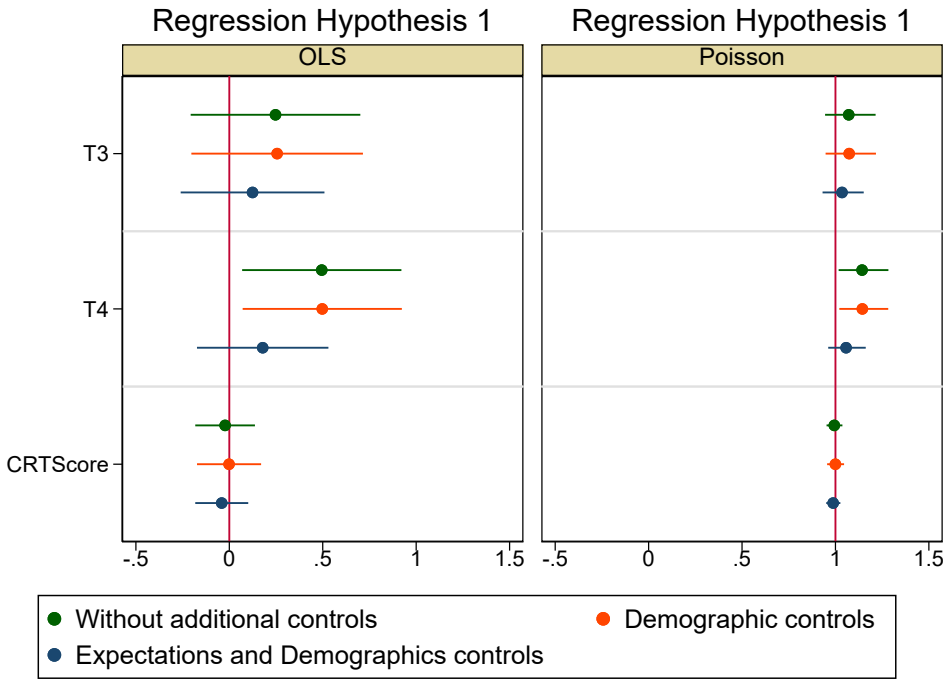

Figure C12: Regression coefficients Hypothesis 1 from OLS and Poisson regressions

Extended lines depict 95% confidence intervals. In all regressions the dependent variable was the number of tokens placed in the blue bucket. Poisson regression report incident rate ratios.

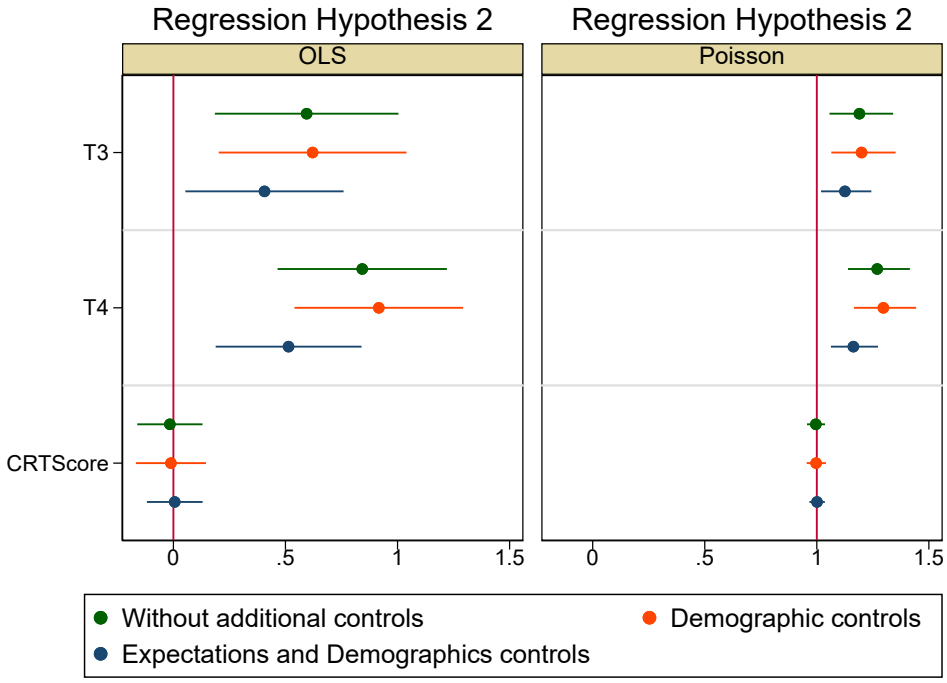

Figure C13: Regression coefficients Hypothesis 2 from OLS and Poisson regressions

Extended lines depict 95% confidence intervals. In all regressions the dependent variable was the number of tokens placed in the blue bucket. Poisson regression report incident rate ratios.

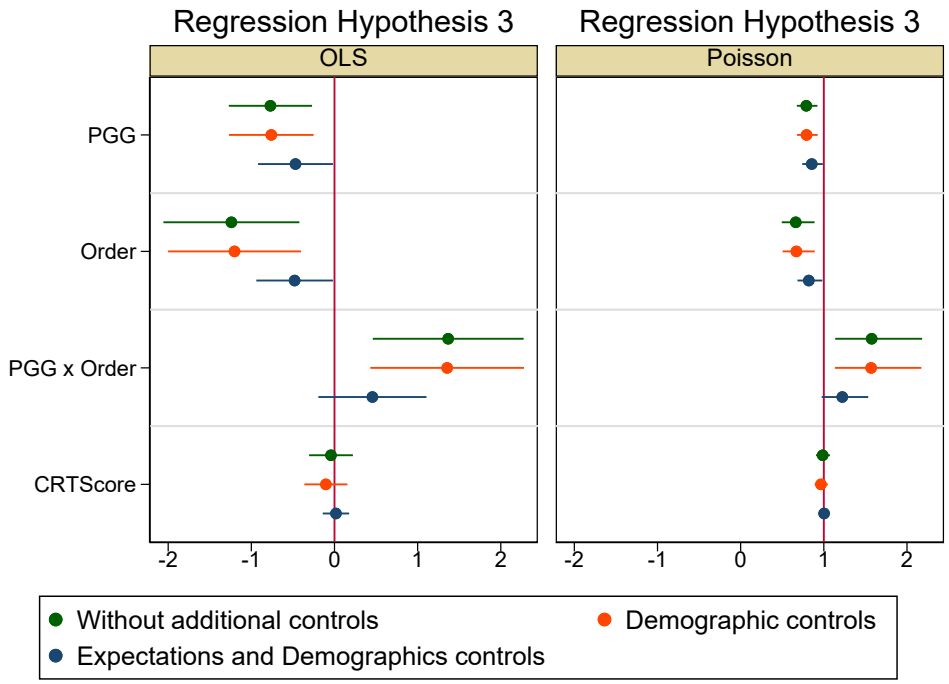

Figure C14: Regression coefficients Hypothesis 3 from OLS and Poisson regressions

Extended lines depict 95% confidence intervals. In all regressions the dependent variable was the number of tokens placed in the blue bucket. Poisson regression report incident rate ratios.

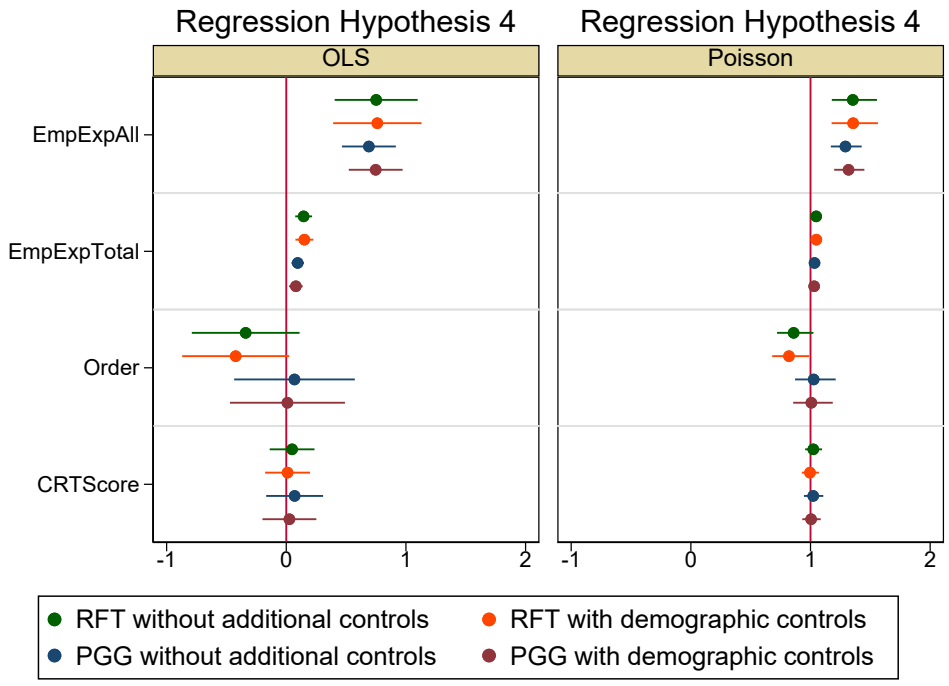

Figure C15: Regression coefficients Hypothesis 4 from OLS and Poisson regressions for T5 (RFT and PGG)

Extended lines depict 95% confidence intervals. In all regressions the dependent variable was the number of tokens placed in the blue bucket. Poisson regression report incident rate ratios.

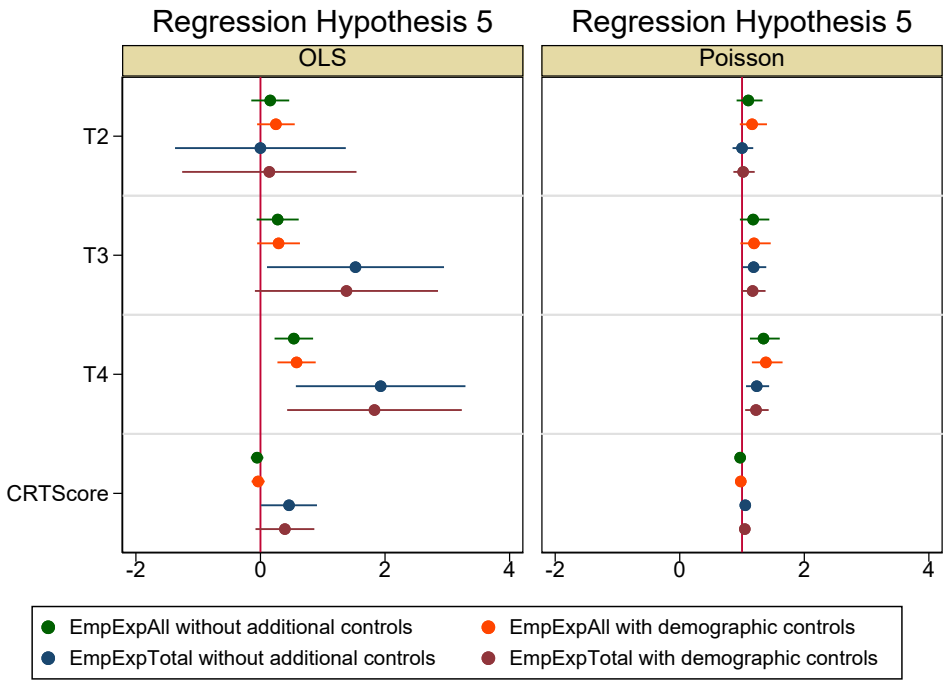

Figure C16: Regression coefficients Hypothesis 5 from OLS and Poisson regressions for T1 to T4

Extended lines depict 95% confidence intervals. In the regressions the dependent variable was either EmpExpAll or EmpExpTotal. Poisson regression report incident rate ratios.

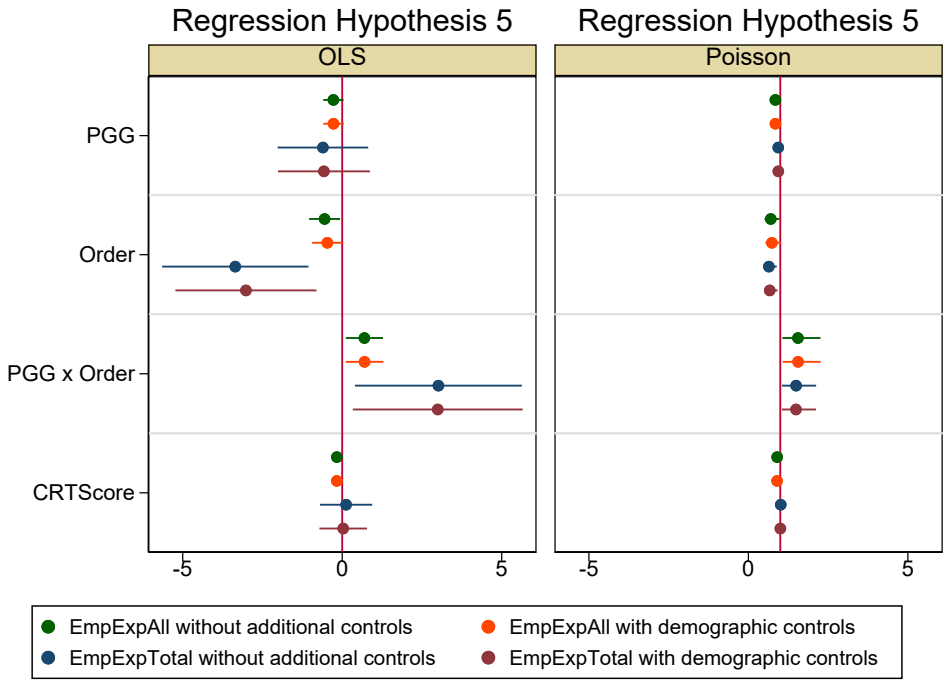

Figure C17: Regression coefficients Hypothesis 5 from OLS and Poisson regressions for T5

Extended lines depict 95% confidence intervals. In the regressions the dependent variable was either EmpExpAll or EmpExpTotal. Poisson regression report incident rate ratios.