Rules, Cooperation and Punishment: Rules in Repeated Public Goods Games

Jack Roycroft-Sherry

August 2025

Abstract: The tendency for people to follow arbitrary rules, even when costly, has been repeatedly demonstrated in experimental economics. Rule following has been proposed as a key reason why humans are capable of sustaining long-term cooperation. However, the mechanisms by which rule-following leads to cooperation are not well understood. This paper aims to shed light on how rules govern cooperation by studying a repeated public goods game with an explicit rule prescribing contributions go to the shared public good, compared to when the rule is absent. Moreover, treatments with peer punishment are also considered, as punishment has been proposed as a crucial element for long-run cooperation. The motivation is that such a design will be able to shed light on how a rule governs behaviour in repeated interactions: for example, how social expectations are shaped by rules, how individuals of different rule-following propensities change their behaviour in the presence of one another, and how the efficacy of punishment is altered when a rule is present.

1.Introduction

Rules are principles or maxims that prescribe or proscribe a particular standard of behaviour (‘do x!’, ‘don’t do y!’) and are ubiquitous in society. For example, they can come as orders (e.g., do not walk on the grass), regulations and guidelines issued by authorities (e.g., tax regulations that require voluntary reporting for some sources of income), as laws and legal statutes (e.g., speed limits on a motorway), and as informal social and moral norms (e.g., forming an orderly queue, or what clothes you can wear) (Gächter et al., 2025).

That people follow rules is becoming increasingly established in experimental economics. Kimbrough and Vostroknutov (2016) find that 62.5% of people follow a rule indicated by a traffic light task where they could earn more money by not following the rule and where there were no consequences if they failed to comply. Gächter et al. (2025) replicate this respect for rules, finding a high 58% rule-conformity rate even with stringent comprehension questions. Moreover, Gächter et al. (2025) study why people follow rules in a systematic way, developing the CRISP framework, which decomposes rule-following into intrinsic respect for rules, material incentives, social expectations (normative and descriptive beliefs about others' rule-following behaviour), and social preferences.

Rules have been proposed as a key reason for widespread cooperation and order in society (Bicchieri, 2006; Columbus et al., 2023; Gächter et al., 2025). Rules may facilitate cooperation by, for example, acting as focal points for coordination (Galbiati & Vertova, 2014), signalling norms of what actions are appropriate and inappropriate (Kimbrough & Vostroknutov, 2016), and highlighting what incentives exist for reward and punishment (Gächter et al., 2025). However, while rule-following is being increasingly established and is proposed to play a crucial role in sustaining cooperation, there remains much to be understood regarding how rules play such a role, particularly in repeated settings.

The public goods game is a setup in experimental economics that has been frequently used to study cooperation (see Ledyard, 1995; Chaudhuri, 2011 for reviews). Galbiati and Vertova (2014) implement non-binding rules in a one-shot public goods game, showing that rules requiring certain levels of endowment contributions (20% or 80%) increase contributions, so long as the rule requires a high enough level (the difference was only significant from no rule in the treatment with the 80% rule, not 20%). More recent work by Gately (2025, forthcoming) also shows that even though a rule is not binding, it nevertheless raises contributions, in particular by increasing unconditional contributions from individuals with a high propensity for rule-following.

However, repeated public goods games, which have different dynamics than one-shot settings, have not been studied systematically with rules. The literature on repeated PGGs shows cooperation to be difficult to sustain, as evidenced by the established finding of declining contributions over rounds (Fehr & Gächter, 2000; Chaudhuri, 2011; Isaac & Walker, 1988). A repeated setting is necessary to examine the interaction between the norm of conditional cooperation and the different preferences and beliefs of individuals (Fischbacher, Gächter, & Fehr, 2001; Kimbrough & Vostroknutov, 2016). A repeated setting with a rule would be particularly interesting for understanding the interaction between individuals with different rule-following propensities (Gächter et al., 2025; Kimbrough & Vostroknutov, 2016), since those with a higher propensity for rule-following contribute differently when a rule is present (Gately, 2025). For example, I hypothesise that a rule will increase the variance of group contributions, as outcomes become more contingent on the individuals present and their past behaviour (see Hypotheses section for more details).

This paper proposes a set of experiments in a registered report style (including pre-registered hypotheses to be tested) to understand the impact of rules in repeated public goods games. I propose four between-subject treatments of 20-round repeated PGGs. The first set of treatments (1 & 2) will have no punishment; Treatment 1 will have no rule and Treatment 2 will have a rule. Treatments 3 and 4 will both have punishment, with Treatment 3 having no rule and Treatment 4 having a rule.

The reason for studying peer punishment is that it has been shown to be a powerful technology for raising contributions. However, punishment can be costly, and it takes time to recoup the losses it inflicts (Herrmann, Thöni, & Gächter, 2008; Gächter, Renner, & Sefton, 2008). Yet given people’s tendency to abide by rules even when individually costly, rules may offer a lower-cost alternative for cooperation (Gächter et al., 2025). Moreover, the combination of rules and punishment could make punishment more efficacious by helping coordinate contributions and who to punish, making it less costly and more efficient. When a rule and peer punishment are combined, I hypothesize that contributions will be higher, and that punishment will be more effective—needing to be used less often and less antisocially—which would increase the earnings efficiency of participants (see hypotheses section).

In combination with these four primary treatments, a week beforehand I will measure individual differences in rule-following and collect other data. The rule-following measure is important for understanding how different types of individuals condition their behaviour on one another. Other measures will include social norms (using the Krupka Weber (2013) norm elicitation method) and conditional preferences (using the strategy method), as measuring these will be important to see how people perceive the public goods game situation and how that predicts their behaviour. I will also use free-form responses and elicit beliefs about others' contributions every five rounds to better understand how and why individuals contribute the way they do over time.

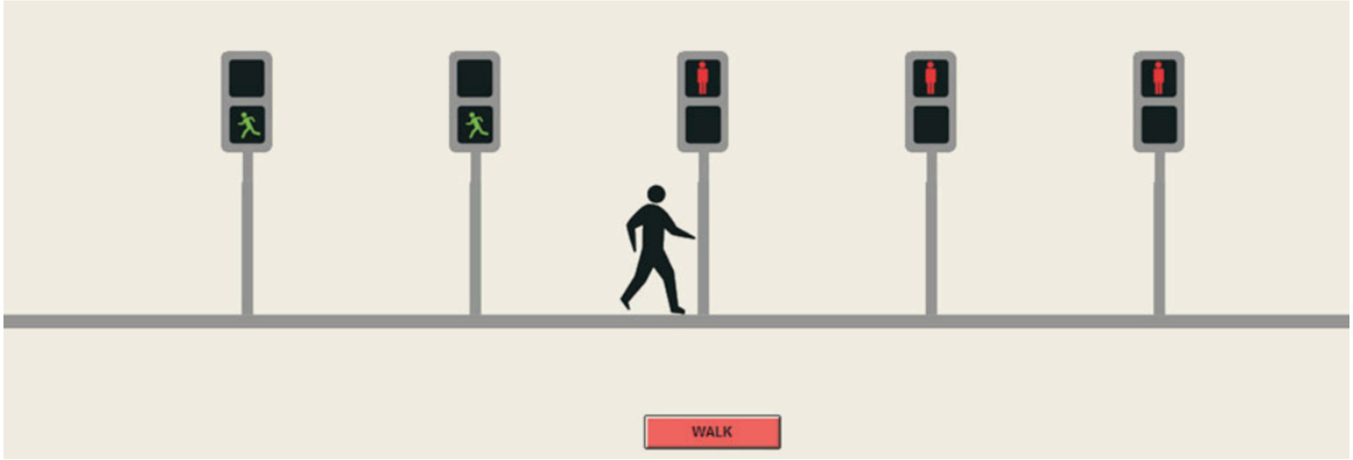
This paper is structured as follows. The literature section details rule-following in experimental economics, cooperation in repeated PGGs, and the interaction between PGGs and explicit rules. The design section details the pre-experiment measurements, PGG parameters, proposed sample size and other design considerations. The hypotheses section pre-registers the hypotheses to be tested. This is followed by an evaluation of the design choices and a discussion of future work. An appendix with proposed experimental instructions is also included.

2. Literature

2.1 Rule Following

Kimbrough and Vostroknutov (2016) helped establish that individuals will often follow a rule even when it is costly to do so and when they could gain financially by breaking it. They demonstrated this using a rule-following traffic light task. In the treatment where a rule was present, participants were instructed to proceed only when the light turned green. However, participants began with a financial endowment that diminished the longer they waited at the initial red light, creating a monetary incentive to violate the rule. They found that 62.5% of people waited at all the traffic lights in the rule condition compared to only 12.5% in the no-rule condition, establishing significant rule-following even when there was no material reason to do so and money could be gained by not complying.

Figure 1: Traffic light task in Kimbrough and Vostroknutov (2016)



Their work shows that rule-following exists on a spectrum, with some individuals following the rule completely, others violating it to different degrees, and some violating it entirely. They argue that this task measures an individual's rule-following propensity, which predicts their sensitivity to the prevailing norm of a given situation. This norm can vary, for example, from conditional cooperation in public goods games to equal splits in ultimatum games.

They provide evidence for this by using their measure of rule-following to predict conformity with the norm in different experimental games. Most relevant to the present paper, in a subsequent public goods game (PGG), individuals were unknowingly sorted into groups based on their rule-following (RF) scores. Groups of high rule following propensity sustained cooperation over 10 periods. Notably, however, both mixed groups and groups composed entirely of low rule-followers exhibited the typical pattern of decaying contributions and were not significantly different from one another. In the mixed groups, the presence of low rule-following types who did not contribute conditionally caused overall group contributions to decline, as the rule-followers adjusted their behaviour in response to others' lack of contribution. They argue that it is not that rule-following types have greater social preferences, but that they more strictly follow the norm of conditional cooperation. Thus, group composition is the key determinant of cooperation levels.

This paper is crucial for establishing the importance of respect for rules, individual differences in adherence to them, and how that manifests in scenarios such as the public goods game. My proposed experiment will expand upon this work by investigating the effect of an explicit rule within a repeated PGG. The setup of the mixed groups in Kimbrough and Vostroknutov (2016) is analogous to Treatment 1 ("no rule, no punishment") in my proposed design, as I will be randomizing individuals into groups rather than sorting them by rule-following propensity. In such a condition, contributions are expected to decline over time. This will then be compared to Treatment 2, which introduces an explicit rule prescribing contributions to the shared public pot.

A more recent paper by Gächter et al. (2025) replicates the finding that many people follow rules even when material incentives suggest they should not, doing so in a traffic light task (where 58% of people followed the rule even with stringent controls) and also in a more abstract task. It then extends work on rule-following to understand why people follow rules, not just that they do.

Gächter et al. (2025) propose a framework that posits rule-following is a function of intrinsic respect for rules, social expectations, incentives, and social preferences (CRISP). The study uses a set of different experiments to control for these various aspects. While all of these factors are potentially important, intrinsic respect for rules is presented as particularly surprising and significant, remaining even when all other factors are absent.

For example, Gächter et al. (2025) look at how people change their rule-following behaviour depending on the appropriateness of following the rule (using the Krupka and Weber (2013) method for eliciting social norms) and their beliefs about whether others follow the rule. Rule-following decreased from 56% (when 80–100% of others disapproved of rule violations) to 35% (when only 0–20% disapproved), and from 56% (when 80–100% of others complied to the rule) to 28% (when only 0–20% complied). Therefore, while people do respond to social expectations and beliefs by reducing their rule-following, a significant level of compliance remains. This helps establish a more refined concept of intrinsic respect for rules, which persists even when there are very few social reasons to follow them. This is relevant to my design because social expectations may change in the presence of a rule and also over time based on the past behaviour of others. By collecting data a week beforehand on individuals' rule-following behaviour, it will be possible to see how different types of individuals condition their behaviour on one another when playing in groups.

Another relevant experiment in Gächter et al. (2025) involves participants completing a rule-following task after observing peers who either follow or violate the rule. The study finds that the presence of just one peer who violates the rule causes rule-following to drop by 8 percentage points. When all peers violated the rule, rule-following dropped from a baseline of 77% to 55%. While people adjust their behaviour in response to peer violations, rule-following remains remarkably high. This is relevant to a repeated public goods game, as it raises the question of how rule-followers will react when some people fail to contribute. It suggests that while they may decrease their contributions, they may not abandon the rule entirely. Therefore, it will be interesting to see how and to what degree different types of individuals adhere to a contribution rule when interacting with one another.

Overall, Gächter et al. (2025) provide a systematic investigation into rule-following and the various elements that can influence this behaviour. When rules are introduced into a public goods game, they can alter social expectations and beliefs, which have been shown to determine contribution behaviour (Kölle, 2015). It is therefore important to see how this plays out in a public goods game setting. That is why in my experimental design, I focus on taking measurements of individual rule-following propensity, perceived norms of the different games beforehand, and measurements of beliefs and free-form responses during the game, to understand the reasons for people's contribution behaviour.

Finally, Gross and De Dreu (2020) sought to understand how group composition affects lying, which was previously proposed to be related to rule-breaking. They first confirmed that individuals classified as rule-followers were less likely to lie in a standard die-rolling task (Fischbacher & Föllmi-Heusi, 2013). The main aim, however, was to see how groups with different compositions of rule-followers and rule-violators behave. In their experiment, groups of four decided on a die-roll number to report, with higher numbers yielding greater monetary rewards. They found that groups composed entirely of rule-violators reported significantly higher numbers (i.e., lied more) than groups composed entirely of rule-followers.

Crucially, the presence of a single rule-follower in a group of rule-violators significantly reduced the amount of lying. Conversely, a single rule-violator in a group of otherwise staunch rule-followers did not increase lying. Gross and De Dreu (2020) argue this is because rule-followers adhere to a norm of honesty and are resolute in their commitment, which in turn influences the group's behaviour.

This raises interesting questions about the prevailing norms when an explicit rule is introduced into a public goods game. As Gately (forthcoming 2025) finds, the introduction of a rule makes contributing more socially appropriate. Therefore, the operative norm may shift from conditional cooperation to one of full contribution. Will rule-followers adhere to this stricter norm even in the presence of rule-violating types?

2.2 Cooperation, Public Goods Games, and Punishment

This section reviews the literature on public goods games (PGGs), a key tool for understanding how cooperation occurs and is sustained. A robust finding in this area is that contributions in standard repeated PGGs tend to decline over time (Chaudhuri, 2011; Isaac & Walker, 1988). However, Fehr and Gächter (2000) demonstrated that introducing costly peer punishment opportunities can reverse this decay, leading to significantly higher and sustained contribution levels. Their experimental design, which had groups play for ten rounds without punishment followed by ten rounds with it (and also the converse), showed that contributions start higher and rise over time when punishment is available. This has led some to propose that punishment is a technology that was evolutionarily necessary for establishing norms and cooperation (e.g., strong reciprocity theory; Fehr, Fischbacher, & Gächter, 2002), as it can effectively increase contributions.

However, research since this foundational paper has provided a more nuanced perspective on the effectiveness of punishment. Its success depends heavily on its cost-benefit structure; punishment is most effective at sustaining cooperation when it is relatively inexpensive for the punisher and sufficiently costly for the punished (Egas & Riedl, 2008; Nikiforakis & Normann, 2008). Furthermore, the potential for anti-social punishment—where cooperators are punished—can undermine its effectiveness (Herrmann, Thöni, & Gächter, 2008). It is also noteworthy that punishment often only becomes profitable for the group over longer time horizons; Gächter, Renner, and Sefton (2008) found that the break-even point, where cumulative earnings in the punishment condition (net of punishment costs) surpass those in the treatment without punishment, is around 17 rounds.

A key contribution of this paper, therefore, is to explore how rules can sustain cooperation, given that peer punishment alone has significant downsides that limit its explanatory power for human cooperation. This study investigates whether the presence of rules can be a factor in sustaining long-term cooperation. I hypothesise that when peer punishment opportunities and a rule are combined, it may enhance the effectiveness of punishment, for example, by clarifying the norm and thus making punishment more impactful on behaviour, by encouraging punishment early on when it is most needed, and by reducing the incidence of anti-social punishment (specific hypotheses will be detailed in the experimental design section). Moreover, a rule alone could potentially increase contributions, a possibility that will be explored later. Therefore, rules may mitigate the need for punishment or even act as a substitute for it.

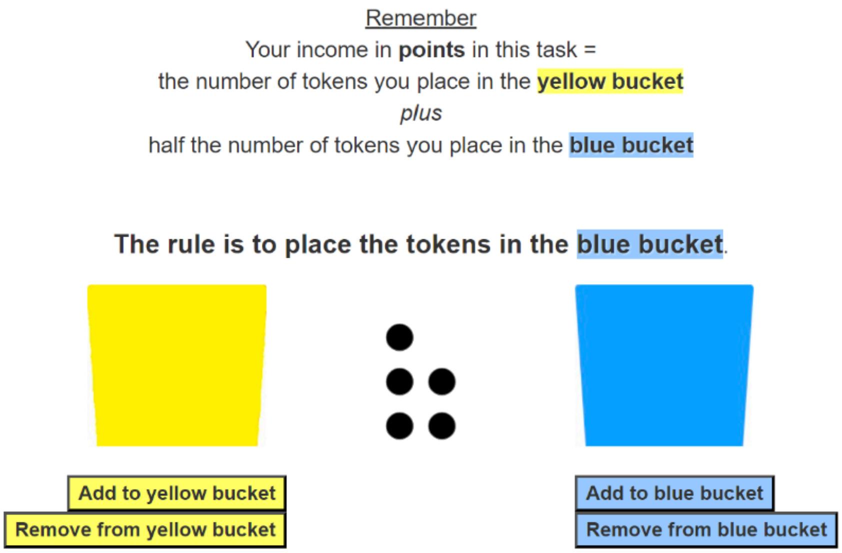
2.3 Rule following and Public Goods Games

One of the earliest studies to bridge rule-following with public goods games comes from Vertova (2014), who examined the influence of "obligations" (i.e., rules) on contributions in a one-shot PGG. The study found that rules can significantly increase contributions, but their effectiveness depends on the standard they set. A high rule, requiring an 80% contribution, significantly increased average contributions compared to a baseline with no rule. However, a low rule requiring only a 20% contribution failed to produce a statistically significant increase. Vertova (2014) also provides evidence as to why a rule has this influence by measuring beliefs and conditional cooperation, finding that the rule works by raising participants' beliefs about what others will contribute and increasing their own willingness to contribute.

Furthermore, Vertova (2014) demonstrated that a rule coupled with even minimal material incentives can substantially boost compliance. When the high rule was paired with a small, probabilistic monetary deduction for non-compliance, contributions rose even higher. This occurred despite the fact that the expected cost of the deduction was so low that a purely self-interested individual should have ignored it. This finding is consistent with Gächter et al. (2025), who also found that a small (10%) probability of an insignificant monetary loss increased adherence to a rule.

Gately (forthcoming 2025) offers a more systematic analysis of how a rule changes the public goods game environment. He uses an innovative design from Kimbrough and Vostroknutov (2018) that measures individual rule-following in one task and then maps the exact same format onto a PGG. The task involves placing tokens into one of two buckets (e.g., yellow or blue), where one bucket provides a higher private payoff. A rule can then be introduced that instructs participants to place tokens in the lower-payoff bucket, pitting self-interest against the rule. This setup can then be adapted into a PGG, where the high-payoff yellow bucket represents a private pot and the blue bucket represents a shared public pot that benefits the group as a whole (see Figures 2 and 3).

Figure 2: Individual rule-following task (Gately, 2025)



Remember

Your income in **points** in this task =

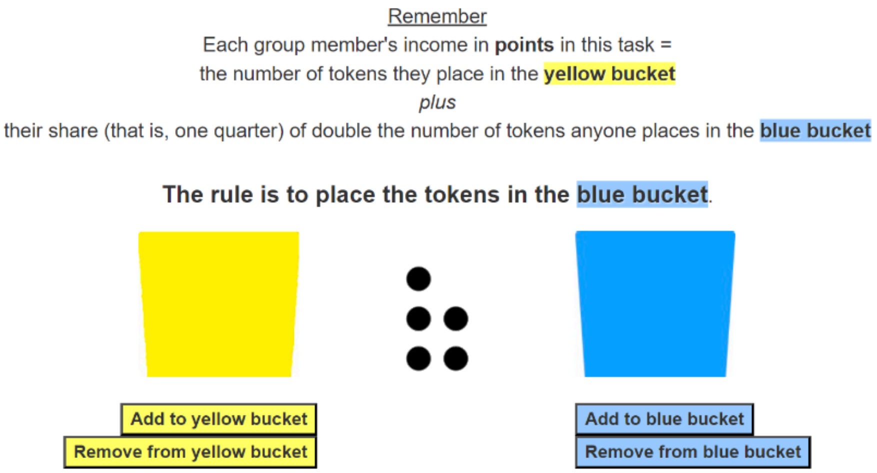
the number of tokens you place in the **yellow bucket**

plus

half the number of tokens you place in the **blue bucket**

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

Figure 3 : Public goods game with rule (Gately, 2025)



Remember

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**

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

In the individual rule-following task, Gately (2025) finds that when a rule is present, 48% of people put all five of their tokens in the designated (but less profitable) blue bucket. This replicates the finding from the broader literature that many people have a high intrinsic motivation to follow rules. Next, in the PGG with the same ball-and-bucket layout, a rule stating "The rule is to place the tokens in the blue bucket" (the public good) significantly increased average contributions to 3.7 tokens, compared to 3.1 tokens in a standard PGG without a rule. The key finding, however, is that this increase was not driven by a change in conditional cooperation. Instead, the rule prompted a rise in the number of people who contributed all five of their tokens unconditionally to the public good. While not everyone followed the rule completely, it exerted a "pull" on behaviour, raising contributions at every level. Gately shows that this can be explained by the interaction between rule-following propensity and normative expectations (using the Krupka-Weber method) and descriptive beliefs about who else contributes to the shared pot (which rise when a rule is present).

Finally, Columbus (2023) provides a related study in a repeated PGG setting. While this experiment was not primarily designed to study the effect of rules versus no rules, it examines the effect of different enforcement mechanisms (no audit, fair audit, biased audit) and group compositions (homogeneous vs. heterogeneous) given that a rule is already in place. I mention it because it uses a repeated PGG (10 rounds). My experiment will build on this by using a longer, 20-round design and, more importantly, by focusing on the direct comparison between a PGG with a rule and one without, allowing for a precise test of how rules themselves shape cooperation over time.

3. Experimental design

3.1 Overview

The main experiment consists of four between-subjects treatments designed to isolate the effects of an explicit rule and a peer punishment mechanism on cooperation in a repeated public goods game (PGG). The four treatments are:

- Treatment 1 (Baseline): A standard PGG with no rule and no punishment.
- Treatment 2 (Rule Only): A PGG with an explicit rule to contribute, but no punishment.
- Treatment 3 (Punishment Only): A PGG with no rule, but with opportunities for peer punishment.
- Treatment 4 (Rule & Punishment): A PGG with both an explicit rule and peer punishment.

A between-subjects design was chosen to ensure that each participant only plays a single game. This approach prevents potential confounds such as learning effects, strategic carryover, or fatigue that could arise if participants were exposed to multiple, similar conditions. The games will be 20 rounds long to provide a sufficient time horizon to observe important dynamic patterns, such as the potential decay of cooperation or the long-term efficacy of rules and punishment.

The primary analysis will involve comparing Treatment 1 with Treatment 2 (to isolate the effect of the rule) and Treatment 3 with Treatment 4 (to test how a rule modifies the effectiveness of punishment). The specific hypotheses for these comparisons are detailed in the next section.

My target sample size is 120 subjects per treatment (for a total of 480 subjects), which would yield 30 independent groups of four in each condition. This target is based on common practices in the field and is intended to provide sufficient statistical power for detecting meaningful differences in first-round and average contributions. However, this number may be adjusted based on a formal power analysis and the relative importance of testing specific hypotheses, balanced against the practical costs of conducting in-person laboratory experiments.

To understand the mechanisms driving behaviour in the main experiment, a survey will be administered to all participants one week beforehand. This pre-experiment session will measure individual rule-following propensity, normative expectations about contribution levels in the different game conditions, and conditional cooperation preferences using the strategy method. Collecting this data in advance allows for the measurement of expectations without priming participants immediately before they make their decisions in the main experiment. This will enable a deeper analysis of the reasons for observed contribution patterns. The details of these pre-experiment measurements are discussed in the following subsection.

3.2 Pre-Treatment Measurements

A session will be conducted with all participants one week before the main experiment to gather data on individual characteristics. The session is timed one week in advance as a practical compromise; a longer interval could be better so that participants play the public goods games without changing their decisions based on these preliminary measures, which will include an explanation of the public goods game in order to elicit social norms. However, a shorter interval is more feasible for minimizing participant attrition.

Individual Rule-Following Task: To obtain a baseline measure of each individual's intrinsic propensity to follow costly rules, participants will complete a one-shot ball-and-bucket task similar to that in Gately (2025). The rule instructs participants to "place the tokens in the blue bucket," while the incentive structure makes placing them in the yellow bucket privately optimal. Using the same ball-and-bucket setup that will be used in the main PGG experiment serves two purposes: it familiarizes participants with the interface and it also makes behaviour between the individual and group tasks more directly comparable. The resulting data can be used to classify individuals (e.g., via a median split) or as a continuous measure of rule-following (how this will be used is discussed in the Hypotheses section).

Social Norms: The Krupka-Weber (2013) method will be used to elicit social norms for all four treatment conditions. In this incentive-compatible task, participants rate the social appropriateness of different contribution levels. They are paid based on how well their ratings match the modal rating of others, encouraging them to reveal their true perception of the social norm. This measurement is essential for testing the hypothesis from Kimbrough and Vostroknutov (2016) that rule-followers are more sensitive to social norms. To do this, one must first measure what the social norms are in the different settings, as the presence of a rule and/or punishment may change them.

Conditional Cooperation Preferences: Conditional cooperation schedules will be elicited using the strategy method. Participants will indicate how many tokens they would contribute for every possible average contribution level of their group members.

Kimbrough and Vostroknutov (2016) inferred conditional cooperation from in-game play, but the strategy method provides a direct measure of underlying preferences. This is important because observing higher contributions when a rule is present is not sufficient to understand the mechanism; if everyone else is contributing more, a conditional cooperator will also contribute more, making it difficult to disentangle their preference from their beliefs about others' actions. The strategy method allows me to capture their complete preference schedule, revealing what they would have done in the absence of high group contributions. This also allows for a direct test of one channel identified by Galbiati and Vertova (2014), where a rule may increase contributions by making people more conditionally cooperative.

Additional Measures: Demographic information (e.g., age, gender) will be collected from the laboratory's participant database and used as control variables. Additional psychometric scales, such as a measure of moral universalism (Enke, 2023), may also be included to provide further information on relevant individual traits. Measuring these items beforehand reduces the cognitive load on participants during the main, 20-round experiment.

3.3 Public Goods Games Details

The public goods games will use the ball-and-bucket task format adapted by Gately (2025, forthcoming) from Kimbrough and Vostroknutov (2018). This format is chosen for two key reasons. First, Kimbrough and Vostroknutov (2018) show that this abstract task generates more variation in rule-following behaviour compared to other tasks, with fewer participants choosing the extreme options of complete compliance or complete violation, which increases the statistical efficiency of estimates. Second, Gately (2025) has demonstrated that this format can be adapted to a PGG setting.

In the treatments with a rule, the wording will be identical to that in Gately (2025): “The Rule is to place balls into the blue bucket.” While other studies have implemented rules as specific contribution levels (e.g., 20% or 80% of the endowment in Galbiati &

Vertova, 2014; 50% in Columbus, 2023), a simple, non-quantified directive was chosen because it mirrors many real-world rules (e.g., "do not walk on the grass"), where individuals must decide for themselves the extent to which they will conform.

Participants will be kept in the same groups for all 20 rounds (Partners matching). This protocol was chosen because the primary focus of the study is on repeated interactions and the dynamics that unfold over time within a stable group. While a Strangers design (where groups are rematched every round) could also be interesting, particularly for studying the role of rules among anonymous individuals, the Partners design is better suited for examining how behaviour is conditioned on the specific history of interactions within a group.

Public Goods Games Parameters

The specific parameters for the PGG are chosen to be consistent with established literature, facilitating comparison with previous findings.

- **Group Size:** 4 participants per group. This is a standard group size in PGG experiments (e.g., Fehr & Gächter, 2000) and is not so large as to dilute individual impact on the group outcome.
- **Marginal Per Capita Return (MPCR):** 0.5. This value also replicates Gately (2025) and means that any contribution to the public pot is doubled for the group. This 2x multiplier is simple for participants to understand.
- **Number of Rounds:** 20. A 20-round design provides a longer time horizon than the more common 10-round experiments. This is particularly important for the punishment treatments, as it allows sufficient time to observe whether the efficiency gains from punishment can offset its initial costs (Gächter, Renner, & Sefton, 2008). It also balances the need to observe long-term dynamics against the participant burden that increases with each round, especially when collecting additional data like beliefs or free-form responses.
- **Endowment:** 20 tokens per round. This is a standard endowment size used in many influential PGG studies (e.g., Fehr & Gächter, 2000; Herrmann, Thöni, & Gächter, 2008).

- **Punishment Mechanism:** In punishment treatments, participants can spend 1 token to reduce another group member's earnings by 3 tokens. Each participant can use a maximum of 3 punishment tokens on any single member per round, and a maximum of 6 tokens in total per round. This cost-to-impact ratio and structure are very similar to those used in other key punishment studies (e.g., Gächter, Renner, & Sefton, 2008).

Measurements During Treatments

During all treatments, two key measurements will be taken.

Beliefs about Others' Contributions: Participants' beliefs about the average contribution of their three group members will be elicited before the first round (round zero) and then every five rounds thereafter. Understanding participants' beliefs is essential for interpreting their contribution decisions. For example, a low contribution could stem from a lack of cooperative preference or from a belief that others will not cooperate. By measuring beliefs directly, we can better disentangle these motivations and understand the coordinating function of a rule.

Free-form Text Responses: Following the belief elicitation every five rounds, participants will also be asked to briefly answer the question: “In your own words, please tell us the main reason for your contribution decisions over the last five rounds.” This qualitative data will provide valuable insight into the reasoning behind participants' choices, which is particularly important given the novel interaction of a rule in a repeated PGG. The text responses will be analyzed using computational text analysis methods. A primary goal of this analysis will be to track the frequency of specific keywords (e.g., "rule," "fairness," "others") to see how participants' justifications for their behaviour change over time. For example, this will allow for a direct test of whether high contributors initially cite the "rule" as their reason, but later shift their justification to the behaviour of others (e.g., "people are not following the rule" or "others are not contributing").

4. Hypotheses

4.1 PGGs Without Punishment (Treatments 1 and 2)

These hypotheses compare the standard PGG (No Rule, No Punishment) with a PGG that includes an explicit rule to contribute (Rule, No Punishment).

Main Hypotheses

- H1: Initial contributions will be higher in the treatment with a rule.
  - Justification: This prediction is based on evidence from one-shot PGGs where non-binding rules have been shown to significantly increase first-round contributions (Vertova, 2014; Gately, 2025). The mechanism is that individuals with a high intrinsic propensity to follow rules will contribute more from the outset. I will test for a statistically significant difference in first-round contributions between Treatment 1 and Treatment 2.
- H2: Contributions will be sustained at a higher level over time in the treatment with a rule.
  - Justification: In the absence of a rule, contributions are expected to decay over time, consistent with findings from standard repeated PGGs (Kimbrough & Vostroknutov, 2016; Chaudhuri, 2011). In contrast, the rule is expected to act as a coordination device and a normative anchor, helping to maintain higher contribution levels throughout the 20 rounds. I will test whether contributions in the rule treatment are maintained relative to the first round.

Secondary Hypotheses

- H3: The variance of group-level contributions will be higher in the treatment with a rule.
    - Justification: An explicit rule may create greater behavioral divergence between individuals with high versus low rule-following propensities. This could make group outcomes more dependent on the specific composition of types within the group and their reactions to initial contribution levels, leading to higher variance.
- H4: Individuals with higher rule-following propensity will contribute more unconditionally.
    - Justification: This hypothesis aims to replicate the finding from Gately (2025) in a repeated setting. I will use the pre-measured rule-following scores to test if high rule-followers contribute more, particularly in early rounds, regardless of others' behaviour. I will also explore how their behaviour changes over time in response to non-cooperation from others.
- H5: Group composition will have a stronger effect on cooperation in the treatment with a rule.
    - Justification: As found in Kimbrough and Vostroknutov (2016), group composition may not significantly affect the decline of cooperation in a standard PGG. However, with a rule in place, I predict that the number of high rule-followers in a group will be positively correlated with sustained, high levels of cooperation.

4.2 Public Goods Game with Punishment (Treatments 3 and 4)

These hypotheses compare a standard PGG with punishment (No Rule, Punishment) to one that includes both a rule and punishment (Rule, Punishment).

Main Hypotheses

- H6: Initial and average contributions will be higher when a rule is present alongside punishment.
    - Justification: The rule is expected to act as a clear coordinating device, establishing a strong norm of high contribution from the first round, which punishment then reinforces.
- H7: The presence of a rule will increase the overall efficiency of the punishment mechanism.
    - Justification: This will be tested in two ways:
        - H7a (Higher Net Earnings): Total group earnings, net of punishment costs, will be higher in the treatment with a rule.
        - H7b (Earlier Break-Even Point): The round in which the cumulative group earnings surpass the cumulative costs of punishment will occur earlier in the treatment with a rule.

Secondary Hypotheses

- H8: Punishment will be more targeted and less anti-social when a rule is present.
    - Justification: The rule clarifies the cooperative norm, making deviations more salient. This is expected to lead to punishment being more consistently directed at non-contributors and a reduction in anti-social punishment (the punishment of cooperators).
- H9: Punishment will be more effective at changing behaviour when a rule is present.

    - Justification: Because the rule makes the expected standard of behaviour unambiguous, individuals who are punished for deviating are expected to increase their contributions more significantly in subsequent rounds compared to those punished in the no-rule condition.
- H10 (Exploratory): I will conduct exploratory tests on whether individuals with high and low rule-following propensities differ in their use of punishment (e.g., frequency, severity, and targets).

5.Evaluation and Conclusion

The finding that peer punishment can dramatically increase cooperation in public goods games was a significant contribution to economics (Fehr & Gächter, 2000). Similarly, rules are ubiquitous and increasingly recognized as a crucial element of human cooperation. This study aims to bridge these two important literatures by systematically investigating the effect of an explicit rule in a repeated public goods game, both with and without a punishment mechanism. The goal is to build a broader understanding of how rules, cooperation, and punishment interact over time.

The experimental design focuses on long-term interactions and the mechanisms that drive individual behaviour. By including a series of pre-experiment measurements and in-game belief elicitations, the design is intended to reveal why people behave as they do and how they condition their actions on the behaviour of others. The core of the study remains a simple and clean comparison across four treatments, allowing for clear causal inferences about the influence of rules on cooperation, both in the presence and absence of punishment.

Several alternative designs were considered. One option was to introduce the rule midway through the experiment. This would have allowed for a direct, within-group comparison of behaviour before and after the rule's implementation, which is a powerful way to measure its immediate impact. Such a design would be well-suited to studying how the sudden introduction of a rule changes behaviour, much like a new sign ("do not walk on the grass") might alter public conduct. However, the current between-subjects design was chosen because its primary goal is different: to understand the stable, long-term dynamics of cooperation within an established institutional environment where a rule is either present or absent from the start.

Another alternative was to use a two-player Prisoner's Dilemma instead of a four-player PGG. This would have simplified the group composition into three distinct types (two rule-followers, two rule-violators, or a mixed pair), allowing for a very direct test of how a rule-follower conditions their behaviour on a rule-violator. This design was rejected for two main reasons. First, cooperation in a dyad can be sustained through direct reciprocity ("tit-for-tat"), which complicates the interpretation of the rule's effect. Second, rules are often a group-level phenomenon, and their function as a coordinating device is more salient and relevant in a group setting than in a pair.

In conclusion, there is much left to learn about how rules facilitate cooperation, the conditions under which they fail, and how their effectiveness is shaped by group composition and other institutional factors. This study hopes to contribute to the existing literature by providing a systematic investigation of rules in a repeated public goods setting, laying the groundwork for future research on this important topic.

Bibliography

Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, 14(1), 47–83.

Columbus, S., Feld, L. P., Kasper, M., & Rablen, M. D. (2023). *Behavioural Responses to Unfair Institutions: Experimental Evidence on Rule Compliance, Norm Polarisation, and Trust*. CESifo Working Paper No. 10565.

Enke, B., Rodríguez-Padilla, R., & Zimmermann, F. (2023). Moral Universalism and the Structure of Ideology. *The Review of Economic Studies*, 90(5), 2339-2378.

Fehr, E., & Gächter, S. (2000). Cooperation and Punishment in Public Goods Experiments. *American Economic Review*, 90(4), 980–994.

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

Gächter, S., Gately, P., & Cubitt, R. (2025, forthcoming). *A CRISP Analysis of Rule Following and Cooperation - 3 Studies*.

Gächter, S., Molleman, L., & Nosenzo, D. (2025). Why people follow rules. *Nature Human Behaviour*, 9, 1342–1354.

Gächter, S., Renner, E., & Sefton, M. (2008). The long-run benefits of punishment. *Science*, 322(5907), 1510–1510.

Galbiati, R., & Vertova, P. (2014). How laws affect behavior: Obligations, incentives and cooperative behavior. *International Review of Law and Economics*, 38, 48-57.

Gross, J., & De Dreu, C. K. W. (2020). Rule Following Mitigates Collaborative Cheating and Facilitates the Spreading of Honesty Within Groups. *Psychological Science*, 31(7), 840-849.

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial Punishment Across Societies. *Science*, 319(5868), 1362–1367.

Isaac, R. M., & Walker, J. M. (1988). Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism. *The Quarterly Journal of Economics*, 103(1), 179–199.

Kölle, F. (2015). The ABC of Cooperation in Voluntary Contribution and Common Pool Extraction Games. *Experimental Economics*, 18(3), 431–448.

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

Ledyard, J. O. (1995). Public Goods: A Survey of Experimental Research. In J. H. Kagel & A. E. Roth (Eds.), *The Handbook of Experimental Economics* (pp. 111–194). Princeton University Press.

Appendix

Experimental Instructions

Preliminary Instructions (All Treatments)

Welcome to the experiment. Please read these instructions carefully. The money you earn will depend on your decisions and the decisions of other participants. All your decisions will be anonymous. At the end of the experiment, your total earnings in tokens will be converted to pounds at a rate of **1 token = £0.01**, and you will be paid in cash.

The experiment consists of **20 periods**. In each period, you will be in a group of four participants. This means you will be in a group with **three** other participants. The composition of your group will remain the same for all 20 periods.

Each period is divided into one or two stages.

Stage A: Contribution Stage (All Treatments)

At the beginning of each period, you will be endowed with **20 tokens**. You must decide how to divide these tokens between a private account (the "Yellow Bucket") and a group project (the "Blue Bucket").

- Tokens you place in the **Yellow Bucket** are kept for yourself.
- Tokens you place in the **Blue Bucket** are contributed to the group project.

All four members of your group will make this decision simultaneously.

The Group Project All tokens contributed to the group project (i.e., placed in the Blue Bucket) by all four group members will be summed up and then **doubled**. This total amount will be divided equally among all four group members, regardless of how much each person individually contributed.

Calculating Your Earnings from Stage A Your earnings for the period are calculated as the number of tokens you kept for yourself plus your share from the group project.

Your earnings can be written as:

(Tokens you kept in Yellow Bucket) + (Your share of the Group Project)

This is equivalent to the formula:

Your Earnings=(20−Your Contribution to Blue Bucket)+42×(Total Contributions to Blue Bucket)
Your Earnings=(20−Your Contribution)+0.5×(Total Group Contribution)

For Treatments with a Rule: *(This text block would be inserted here for the relevant treatments.)*

The Rule In this experiment, there is a rule of conduct: **You should place all your tokens in the Blue Bucket.**

Stage B: Punishment Stage (For Punishment Treatments Only)

After the Contribution Stage, you will be shown the individual contributions of each of the other three members of your group to the project for that period.

You will then have the opportunity to reduce the earnings of any of the other group members by assigning them deduction points. The other group members will also have the opportunity to reduce your earnings.

- You can assign between **0 and 3 deduction points** to each of the other three group members.
- You have a total of **6 deduction points** that you can assign in each period.

For each deduction point you assign to a group member, their earnings for that period will be reduced by **3 tokens**.

- Assigning 0 points does not change their earnings.
- Assigning 1 point reduces their earnings by 3 tokens.
- Assigning 2 points reduces their earnings by 6 tokens.
- Assigning 3 points reduces their earnings by 9 tokens.

A participant's total reduction in earnings is the sum of the reductions from all other group members. For example, if two other members assign you 1 deduction point each, and the third member assigns you 2 deduction points, your total deduction points received are 1+1+2=4. Your earnings for the period would then be reduced by 4×3=12 tokens.