Efficiency of the Minimum Approval Mechanism with heterogeneous players

Gabriel Bayle, Marc Willinger

CEE-M
CENTER FOR ENVIRONMENTAL ECONOMICS – MONTPELLIER

Efficiency of the Minimum Approval Mechanism with heterogeneous players

Gabriel Bayle
&
Marc Willinger

CEE-M Working Paper 2025-02

CEE-M

INRAE

cnrs

L'INSTITUT agro Montpellier

Efficiency of the Minimum Approval Mechanism with heterogeneous players

Gabriel Bayle[1,*] and Marc Willinger[2]

[1]GATE, CNRS, Université Lumière Lyon 2
[2]CEE-M, Univ. Montpellier, CNRS, INRAe, Institut Agro
[*]Corresponding author: gabriel.bayle.econ@gmail.com

27th February, 2025

Abstract

The Minimum Approval Mechanism (MAM) was introduced by Masuda et al. (2014) as a mechanism aimed at mitigating free riding in the social dilemma context of a public good game. The MAM is a two-stage mechanism which theoretically achieves the socially optimum level of public good provision, according to various equilibrium concepts (e.g., backward elimination of weakly dominated strategies, level-k, or minimax regret). We study the robustness of this mechanism to the introduction of endowment heterogeneity. We explore, theoretically and experimentally, how endowment inequalities affect the effectiveness of the MAM at improving the level of provision. We find that the mechanism is still Pareto-improving under endowment heterogeneity, but that its efficiency diminishes as inequality is increased. Our experimental findings indicate a significant weakening of the mechanism under endowment inequalities, surpassing our theoretical predictions. A close examination of individual behaviors reveals a significant drop in contributions compared to the uniform case, prompted by even minor inequalities. Intriguingly, our findings challenge conventional assumptions by showing that inequality aversion drives contributions in a public good game with endowment disparities only under certain assumptions. We explore the impact of beliefs about the contributions of advantaged player as potential motivations through guilt aversion and Kantian preferences.

Keywords: Approval mechanism; Public Goods; Inequalities; Inequality aversion

JEL Codes: C72, C92, H41, D9, P43

1 Introduction

Public goods provision is a central theme in both economic theory and policy-making, posing significant challenges due to the free-rider problem and the difficulty of achieving Pareto efficient provision. While it has been largely demonstrated experimentally that, even without regulation, a positive amount of provision is naturally achieved, attaining the social optimum remains a challenge (for comprehensive literature reviews, see Ledyard, 1995; Chaudhuri, 2011). This failure of cooperation is primarily due to an incentive problem that pits selfish interests against collective interests. Several institutional designs have been studied to address this issue, including compensation (Groves, 1973; Laussel and Palfrey, 2003), taxation and allocation (Groves and Ledyard, 1977; Chen and Plott, 1996), and rewarding and punishment (Falkinger, 1996; Falkinger et al., 2000). Two-stage mechanisms are less common, but notable examples include the compensation mechanism by Varian (1994), the second-stage punishment (Fehr and Gächter, 2000), and the approval mechanism initially introduced by Saijo et al. (2011) for Prisoner's Dilemmas and later extended to public goods by Masuda et al. (2014) as the Minimum Approval Mechanism (MAM).

The Minimum Approval Mechanism (MAM) represents an interesting approach to addressing challenges related to public goods provision. The mechanism operates in two stages. In the first stage, all players simultaneously propose an individual level of contribution to the public good. In the second stage, the proposal vector is publicly revealed, and players vote to approve or disapprove it. In case of approval, the first stage proposed contributions are implemented. Otherwise, i.e. in case of disapproval, the minimum contribution level proposed in stage 1 is implemented for all players.

Masuda et al. (2014)demonstrated that in a two-player public goods game, the MAM leads to efficient provision of the public good, as supported by both theoretical analysis and experimental findings. The theoretical result holds across multiple solution concepts, backward elimination of weakly dominated strategies (BEWDS; Kalai, 1981), subgame perfect minimax regret (Renou and Schlag, 2011), limit logit agent quantal response equilibrium (limit LAQRE; McKelvey and Palfrey, 1998) and two heuristics, Level-K (Nagel, 1995) and Diagonalization. Experimentally the results have been as convincing by showing a quick convergence to the socially optimal provision. Masuda et al. (2014) result was

extended to two-player common pool resource (CPR) games by Yao et al. (2022). The extension of MAM to three-player Common-Pool Resource (CPR) games (Yao et al., 2024a) and three-player public goods games (Yao et al., 2024b) demonstrated similar efficiency outcomes when unanimity approval was achieved. However, under majority approval, the mechanism did not consistently yield efficient outcomes. This limitation highlights challenges in generalizing the effectiveness of the MAM across different approval rules and game structures.

In this paper, we investigate the impact of heterogeneity on the efficiency of the mechanism, focusing on endowment asymmetry within the context of public goods provision. This choice of heterogeneity enables us to examine closely the dynamics of cooperative behavior under conditions that mirror real-world economic disparities, including factors like exposure to climate change and efforts to reduce CO2 emissions. Specifically, our study examines the hypothesis that while income inequality may weaken the mechanism, it should still lead to Pareto improvements. Additionally, we aim to investigate how other-regarding preferences affect contributions, with and without the mechanism. Building on insights from Fehr and Schmidt (1999), we consider how heterogeneous preferences concerning advantageous and disadvantageous inequalities, affect contributions and voting decisions in context of the MAM.

To evaluate the effectiveness of the Minimum Approval Mechanism (MAM) in the presence of endowment inequality, we conducted a laboratory experiment using a between-subjects design. Participants were randomly assigned to one of several conditions, each representing a different level of endowment inequality (low, high, or no inequality), either with or without the MAM. In each condition, participants played a repeated public goods game over 19 rounds. Throughout the experiment, players were consistently assigned to either a "rich" or "poor" role. with players assigned to either a "rich" or "poor" role for the entire experiment. The "no inequality" treatment was included to replicate the original findings of Masuda et al. (2014). Additionally, we measured participants' inequality aversion using dictator and ultimatum tasks to quantify their sensitivity to advantageous and disadvantageous inequalities. We also gathered data on participants' beliefs about their partners' expected contributions to assess the influence of social expectations on decision-making. This comprehensive dataset allowed us to analyze the impact of endowment inequalities

on the effectiveness of the MAM and to determine whether observed behaviors matched theoretical predictions based on inequality aversion.

We observe larger contributions with the Minimum Approval Mechanism (MAM) compared to without it. While the MAM leads to Pareto improvements experimentally, these improvements are less pronounced than theoretical predictions suggest. Surprisingly, inequality aversion does poorly predict contribution and voting behaviors. However, both the rich and the poor players share the common belief that richer players should contribute more. Nevertheless, the beliefs of both types reveal that the expected contributions are lower than those predicted.

The rest of the paper is organized as follows. In Section 2 we present our theoretical predictions regarding the influence of income inequality on the effectiveness of the MAM, and about the impact of inequality aversion preferences. Section 3 describes our experimental design. Section 4 presents the experimental results and Section 4.4 the exploratory results. Section 5 discusses the main findings and concludes.

2 Theoretical framework

2.1 Standard model

Consider a two-player Public Good (PG) game. Each player is endowed with $w_i > 0$ units of cash, which they must allocate between a private activity and their contribution to the PG. Let g_i denote the contribution of player i, where $i = 1, 2$, and $0 \leq g_i \leq w_i$. Let $\pi_i(g_1, g_2)$ denote the corresponding level of profit. Define $G = g_1 + g_2$ as the aggregate level of contribution. The profit for player i is given by:

$$\begin{aligned} \pi_i(g_1, g_2) &= R_i(g_1, g_2) + Z_i(w_i - g_i) \\ &= \gamma(g_1 + g_2) + w_i - g_i \end{aligned} \tag{1}$$

with γ representing the marginal per capita return (MPCR). We assume that R_i, the payoff from the PG, and Z_i, the payoff function from the private activity, are linear functions.

Under endowment equality, i.e. $w_1 = w_2 = w$, the dominant strategy in the Nash equi-

librium is $g_1^* = g_2^* = 0$, indicating zero provision of the public good. If $\gamma < \frac{1}{2}$, the Nash equilibrium is also a social optimum, meaning there is no social dilemma. Conversely, if $1 > \gamma > \frac{1}{2}$, the socially optimal contribution is $\hat{g}_1 = \hat{g}_2 = w$, resulting in a public good level of $G = 2w$.

In the case of endowment inequality, we assume that player 2 is the richer player, i.e. $w_2 > w_1 \geq 0$, and $w_1 + w_2 = 2w$. For comparability, the group endowment $\sum_{i=1}^{2} w_i = W$ is kept constant. While the Nash equilibrium remains unchanged under endowment inequality, i.e., $g_1^* = g_2^* = 0$, individuals' optimal contributions depend on their endowment, specifically, $\hat{g}_2 = w_2 > \hat{g}_1 = w_1$.

The Minimum Approval Mechanism (MAM) introduces a second stage wherein subjects decide whether to approve (y) or not (n) the vector of proposals $\mathbf{g} = (g_1, g_2)$, with g_1 and g_2 being the contribution proposals of the two players. If both players approve, then $G = g_1 + g_2$. If at least one player disapproves, then $G = 2 \times \min(g_1, g_2)$. Masuda et al. (2014) showed that under the MAM with players with equal endowments, the Nash equilibrium is Pareto optimal with $g^* = w$, and thus $G = 2w$. To study the MAM under heterogeneous player cases, consider $\underline{g} = \min(g_1, g_2)$ and $\overline{g} = \max(g_1, g_2)$. Specifically, by using Backward Elimination of Weakly Dominated Strategies (BEWDS), we first establish two preliminary lemmas, starting with stage 2 approval decisions.

Lemma 1. *Under the MAM, the player proposing the lowest contribution always approves, while the player proposing the highest contribution always disapproves.*

Proof. See appendix A.1.1 □

Note that the intuition is straightforward: the minimum contributor is the one who defines the disapproval benchmark. Therefore, he has an incentive to free ride on the highest contributor's contribution, in other words "approving" is a dominant strategy for him.

Lemma 2. *Under the MAM, only symmetric proposition vectors survive any subgame.*

Proof. See appendix A.1.2 □

In Masuda et al. (2014), the "echelon structure" induced by the MAM plays a key role under BEWDS. Consider any proposal vector (g, g') such that $g < g'$. The contribution

game has an echelon structure if the payoffs for the proposal vectors (g, g') and (g', g) are the same as for vector (g, g) and if the payoff for vector (g', g') is strictly larger than for the vector (g, g). This property follows directly from Lemma 1 and 2. The echelon structure ensures that the highest possible symmetric proposal vector, weakly dominates all other proposal pairs. In the equal endowment case, the highest possible symmetric proposal is the socially optimal contribution, i.e. $g = w$, which leads to $G = 2w$. In our case of endowment inequality, i.e. $w_1 < w_2$, the echelon structure property still applies but only partially, because the highest possible symmetric proposal vector is (w_1, w_1) according to Lemma 2. This is stated as Proposition 1.

Proposition 1. *Under endowment asymmetry, $w_1 < w_2$, the MAM induces a partial echelon structure. The highest payoff occurs for the pair (w_1, w_1), where w_1 is the highest possible symmetric contribution. This partial echelon structure means that contributing $g_i = w_1$ weakly dominates any other strategy for both players.*

Proof. Given any subgame with an asymmetric proposal vector leads to symmetric contributions, we compare $\pi_i(g, g)$ to $\pi_i(g', g')$ for all $g < g'$:

$$\pi_i(g', g') > \pi_i(g, g)$$
$$w_i - g' + 2\gamma g' > w_i - g + 2\gamma g$$
$$g' > g$$

This is true by the assumption. Since w_1 is the maximal contribution that Player 1 can make, under the MAM, $g_2 = w_1$ becomes Player 2's best response if Player 1 contributes $g_1 = w_1$. □

According to Proposition 1, the MAM results in a Pareto improvement, characterized by a symmetric and positive contribution that matches the wealth of the poorest player.

Proposition 2. *Under the MAM, the level of the public good is determined by the wealth of the poorer player, w_1, and increases strictly with w_1.*

Proof. Follows directly from Proposition 1. □

Having examined the strategies under the MAM, we now focus on its impact on inequal-

ities. Does the mechanism influence the initial inequality in endowments? To facilitate this discussion, we will distinguish between ex-ante and ex-post inequalities, recognizing that ex-ante inequality refers to endowment inequality, while ex-post inequality refers to payoff inequality.

This outcome results in both ex-ante and ex-post absolute and relative inequalities. These inequalities are denoted as $\Delta_{a,a}$ and $\Delta_{a,p}$ for absolute inequalities, and $\Delta_{r,a}$ and $\Delta_{r,p}$ for relative inequalities, where the first indices a and r correspond to absolute and relative inequalities, respectively, while the second indices a and p denote ex-ante and ex-post inequalities.

Specifically:

- Absolute inequalities are given by:

$$\Delta_{a,a} = \Delta_{a,p} = w_2 - w_1$$

- Relative inequalities are given by:

$$\Delta_{r,a} = \Delta_{r,p} = \frac{w_2 - w_1}{w_2 + w_1}$$

Proposition 3. *The MAM (i) does not affect absolute inequality since* $\Delta_{a,a} = \Delta_{a,p} = w_2 - w_1$*, but (ii) reduces relative inequality, i.e.,* $\Delta_{r,a} > \Delta_{r,p}$.

Proof. (i) Under the MAM, both players contribute w_1. The resulting payoffs are $\pi_1(w_1, w_1) = 2\gamma w_1$ and $\pi_2(w_1, w_1) = w_2 - w_1 + 2\gamma w_1$. Therefore, the absolute inequality, $\Delta_{a,a}^{MAM} = \Delta_{a,p}^{MAM} = w_2 - w_1$, remains unchanged from the scenario without the MAM to the scenario with the MAM before the public good game.

(ii) The relative inequality without the MAM (or ex-ante with the MAM) is given by the difference in payoffs divided by the total payoff, i.e., $\Delta_{r,a}^{MAM} = \frac{w_2 - w_1}{w_1 + w_2}$. With the MAM, the individual payoff increases by $2\gamma w_1$ for both players, making the total payoff $w_2 - w_1 + 4\gamma w_1$. Hence, the ex-post relative inequality becomes $\Delta_{r,p}^{MAM} = \frac{w_2 - w_1}{w_2 - w_1 + 4\gamma w_1}$. The inequality $\Delta_{r,a}^{MAM} > \Delta_{r,p}^{MAM}$ holds when $\gamma > \frac{1}{2}$, which is true by assumption as $\frac{1}{2} < \gamma < 1$. This analysis implies that the MAM reduces ex-post relative inequalities. □

In our analysis, we initially consider a standard self-centered behavior model, where individuals aim to maximize their own profit. However, given the inherent inequalities present in our setting, it is relevant to study how individuals react to these disparities in endowments and, consequently, payoffs. To do so, we focus on the trade-off between profit maximization, with and without the Minimum Approval Mechanism (MAM), and inequality minimization. Since the outcomes implemented by the MAM do not minimize inequalities (Proposition 3), it is crucial to understand if other outcomes, influenced by preferences for inequality reduction, could lead to a different performance of the MAM. This allows us to explore whether integrating inequality aversion could yield different predictions for the performance of the MAM under varying levels of endowment inequality.

2.2 Inequality aversion model

We now consider a model of heterogeneous agents with inequality aversion preferences *à la* Fehr and Schmidt (1999) in a context of endowment inequality[1]:

$$U_i(g_i, g_{-i}) = \pi_i - \alpha_i \max(\pi_{-i} - \pi_i, 0) - \beta_i max(\pi_i - \pi_{-i}, 0) \tag{2}$$

where $\pi_i = w_i - g_i + \gamma(g_i + g_{-i})$. The parameter α_i represents the aversion to disadvantageous inequality (or "envy"), and β_i represents the aversion to advantageous inequality (or "guilt"), with $\beta_i \leq \alpha_i$ and $0 \leq \beta_i < 1$. Note that in cases of equality, i.e., $\pi_1 = \pi_2$, the outcome is identical to that of selfish agents. This equivalence also holds if $\beta = 0$ and $\alpha = 0$, respectively, for the player with higher profit and for the player with lower profit. It should be emphasized that for the following propositions, we maintain $w_2 > w_1$. To simplify the statement of our key propositions under inequality aversion, we define the "guilt threshold" $1 - \gamma$ for the richer player. We refer to Player 2 as a high guilt type if $\beta_2 > 1 - \gamma$, and as a low guilt type if $\beta_2 \leq 1 - \gamma$.

Proposition 4. *Without the MAM, the Nash equilibrium contribution pair is* $(0, w_2 - w_1)$ *for a high guilt rich player, or the inefficient outcome* $(0, 0)$ *for a low guilt rich player.*

[1]The two types of heterogeneity, in endowments and in preferences, are complementary. If we consider inequality aversion in the original setting Masuda et al. (2014) without unequal endowments, the baseline predictions (see Appendix A.2 for the formal version) are unaffected: only the symmetric contribution vectors survive in every subgame. Therefore, the echelon structure guarantees that the unique equilibrium coincides with the social optimum, regardless of inequality aversion.

Proof. The dominant strategy for player 1 is to contribute $g_1^* = 0$, as $U_1(g_1, 0)$ decreases with g_1 regardless of α_1. For player 2, the contribution depends on her level of guilt aversion, β_2. If $\beta_2 \leq 1 - \gamma$, then $U_2(0, g_2) \leq U_2(0, 0)$, implying player 2's best response, if player 1 contributes 0, is $g_2^* = 0$. Thus, $(0, 0)$ is a Nash equilibrium when the richer player's guilt aversion is low. Conversely, if $\beta_2 > 1 - \gamma$, player 2's dominant strategy is to contribute $w_2 - w_1$, equalizing the players' payoffs when $g_1^* = 0$, minimizing inequality. Note that player 2 will not contribute more than $g_2 = w_2 - (w_1 - g_1)$ to avoid reversing welfare inequality with player 1, which would activate her aversion to disadvantageous inequality, α_2. □

To determine the equilibrium proposal in stage 1, it is necessary to categorize the second-stage approval decisions, as stated in Lemma 3.

Lemma 3. *The second stage approval decisions fall into three categories:*

(i) $g_2 > w_2 - (w_1 - g_1)$*: Player 2 always disapproves.*

(ii) $g_1 < g_2 \leq w_2 - (w_1 - g_1)$*: Player 1 and a high guilt player 2 always approve. A low guilt player 2 disapproves.*

(iii) $g_1 > g_2$*: Player 1 always disapproves.*

Proof. See Appendix A.1.3 □

Proposition 5. *Under the MAM with BEWDS, the equilibrium contribution pair is* (w_1, w_1) *if Player 2 is a low guilt type, and* $(0, w_2 - w_1)$ *if Player 2 is a high guilt type.*

Proof. In the scenario where $\beta_2 > 1-\gamma$ and $g_2 \leq w_2-(w_1-g_1)$, both players approve any vector where $g_1 < g_2$, not leading to symmetric outcomes and echelon structure anymore. Since Player 2's utility increases with g_2, she contributes the maximum amount that does not reverse inequalities, i.e., $w_2 - w_1$. Player 1's optimal response is to contribute $g_1 = 0$.

For $\beta_2 \leq 1 - \gamma$, any asymmetric proposals lead at least one player to disapprove, driving the outcome towards symmetric contributions due to the partial echelon structure. In the normal form game, stage 1, both players will contribute the same amount g. We can write their utility functions: $U_1(g, g) = w_1 - g + 2\gamma g - \alpha_1(w_2 - w_1)$ and $U_2(g, g) = w_2 - g + 2\gamma g - \beta_2(w_2 - w_1)$ both increasing in g when $\gamma > \frac{1}{2}$, which is true by assumption. □

Under endowment inequality, we observe a departure from the standard preference model when rich players are guilt-averse. Specifically, rich players with high levels of guilt aversion tend to contribute more than poorer players, resulting in asymmetric contributions that help mitigate the initial inequality. This suggests that guilt-averse rich players can substitute the approval mechanism by voluntarily increasing their contributions, thereby absorbing the inequality. However, for this outcome to occur, the rich players must exhibit extremely high levels of inequality aversion.

When players show insufficient inequality aversion, the results align with those of the standard model. Under the approval mechanism, both players contribute the maximum symmetric amount. In contrast, in an unregulated public goods game, neither player contributes anything.

3 Experimental design

The study included four treatments, with approximately 80 subjects per treatment, plus 20 subjects to replicate the results of Masuda et al. (2014) that we discuss by the end of this section, totaling 330 subjects. The experiment adopted a between-subject design, whereby each subject underwent only one treatment. Table 1 provides details on the group composition of each treatment. Subjects received an average payment of 21 euros for their participation, including a 5-euro participation fee, with a duration ranging from 1 hour to 1 hour and 30 minutes. At the conclusion of the study, one of the 19 rounds of the public goods game was selected at random, along with one of the elicitation tasks for payment.

Table 1: Group composition of the 6 treatments.

(3x2)	Without MAM	With MAM
Low Inequality	78 subjects	84 subjects
High Inequality	70 subjects	78 subjects
No Inequality	8 subjects	12 subjects

The experiment comprised four tasks and two questionnaires, designed to investigate the effectiveness of the MAM in the context of endowment inequality. The primary task was a repeated Voluntary Contribution to a Public Good game (PG). To ensure comprehension of the PG game, the first questionnaire assessed subjects' understanding of the game mechanics. The secondary tasks aimed to elicit inequality aversion *à la* Fehr and Schmidt

(1999).

The third task was the Number Line Estimation (NLE) test, used to estimate the approximate number system (ANS) (Siegler and Opfer, 2003), which serves as a proxy for numerical and arithmetic understanding. The ANS is an aspect of cognitive function that remains active throughout life, enabling quick and intuitive grasp of numbers and their interconnections (Halberda and Odic, 2015). Additionally, standard socio-demographic information was collected. Finally, participants' perceptions of their financial well-being, both in absolute and relative terms, were recorded.

The experiment was designed using oTree (Chen et al., 2016) and conducted between June and September 2023 at the Laboratory for Experimental Economics of Montpellier (LEEM) with a database of approximately 2000 volunteers, primarily consisting of students from the University of Montpellier. The experimental design[2] and the pre-analysis plan were preregistered on Open Science Framework the May 11th 2023 and validated by the Research Ethics Board of the University of Montpellier the May 23th 2023.

The "No Inequality" treatment replicates the original study, motivated by the variation that our participants were French students instead of Japanese students. To ensure sufficient statistical power, we determined the sample size for this treatment based on a power analysis of the original study. This analysis, considering the large effect size reported in the original study, indicated that this small sample size would provide over 90% power to detect a significant effect at the $\alpha = 0.05$ level. This replication allows us to control for potential sample-related differences between our study and the original. If our main treatments lead to a difference compared to the original study, we can be more confident that these differences are not attributable to variations in sample characteristics. As this study is preliminary and the replication of the original result is confirmed, we describe the results only in a footnote of Section 4.

3.1 Public good game

For the main task, the experiment employed a public good game that was repeated over 19 rounds under (imperfect) stranger matching[3]. Sessions were randomly assigned to one

[2]The instructions are included as online appendix in the supplementary materials in French and English.

[3]Because we ran 19 rounds with a 20-players experimental room, each subject was informed that at most she would play 2 rounds with the same partner.

of the treatments, *i.e.* "with" or "without" the MAM and for a given level of inequality: "Low inequality" with $w_1 = 18$ and $w_2 = 22$, "High inequality" with $w_1 = 10$ and $w_2 = 30$, and "No inequality" with $w_1 = w_2 = 20$. To ensure comparability across the three inequality levels, certain parameters were held constant. Specifically, the per-period group endowment was fixed at $W = w_1 + w_2 = 40$ for all treatments, as was the marginal per capita return ($\gamma = 0.7$). Except for the "No Inequality" treatment, participants were randomly assigned to a player type, with half being designated as "rich" and the other half as "poor". At the beginning of each round, participants within a session were randomly paired with a participant of the other type from the same session. In the baseline treatment, subjects were required to choose a contribution level to the public good. Once chosen, the contribution vector and payoff for the period were announced to both players. They were then matched with new partners for the subsequent round, continuing for 19 rounds. In the approval mechanism treatments, the game was divided into two stages. In stage 1, subjects submitted contribution proposals. In stage 2, the contributions were revealed, and each player was invited to approve or disapprove the proposal vector. If the vector was approved, both players contributed the amount they submitted in stage 1 and proceeded to the next period. If at least one player disapproved the proposal vector, both players contributed the minimum of the proposal vector, and the game continued for the next round with a new partner. Alongside their contribution decision, participants were asked to state what they thought the other player would contribute. This statement was not elicited and allowed the subjects to calculate the outcome and make their own decision.

3.2 Inequality aversion elicitation tasks

For the inequality aversion elicitation task, we followed the experimental design of Blanco et al. (2011), incorporating a Threshold Dictator Game (TDG) and a Strategy Method Ultimatum Game (UG) in a one-shot format. We estimated the parameters α_i and β_i of the inequality aversion model proposed by Fehr and Schmidt (1999) using the methodology outlined in Blanco et al. (2011).

To estimate β_i, which captures advantageous inequality aversion, in the TDG, subjects chose the level of equality, ranging from $(0, 0)$ to $(20, 20)$, for which they were willing to forego the selfish rational outcome $(20, 0)$. We observed the egalitarian outcome (x_i', x_i')

that subject i preferred to $(20, 0)$, considering that they were indifferent between $(20, 0)$ and $(\tilde{x_i}, \tilde{x_i})$ with $\tilde{x_i} \in [x_i' - 1, x_i']$. We then computed β_i as follows:

$$\beta_i = 1 - \frac{\tilde{x_i}}{20} \tag{3}$$

To estimate α_i, which captures disadvantageous inequality aversion, we used the ultimatum game under the strategy method to identify the rejection threshold by considering all distributions from $(0, 20)$ to $(10, 10)$. We observed the approval threshold s_i' and assumed that the indifference level was $s_i \in [s_i' - 1, s_i']$ such that subject i was indifferent between s_i and 0. We then computed α_i as follows:

$$\alpha_i = \frac{s_i}{2(10 - s_i)} \tag{4}$$

We arbitrarily defined $s_i = s_i' - 0.5$ and $\tilde{x_i} = x_i' - 0.5$ as in Blanco et al. (2011).

4 Results

This section is organized as follows. We start by investigating the effectiveness of the MAM under endowment inequality in Section 4.1. In Section 4.2, we discuss the inequality aversion predictions. Finally, in Section 4.3 we focus on the voting behaviors to disentangle role-specific decisions. The replication data of the Masuda et al. (2014)'s setting are excluded from this analysis[4]. In the next Section 4.4, we explore the alternative motivations explaining the observed behaviors.

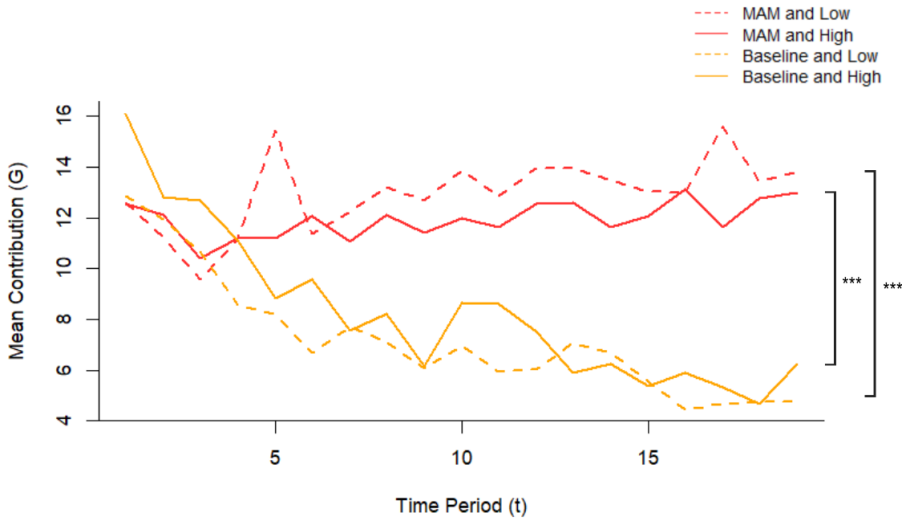
4.1 MAM effectiveness

We first study the effect of the MAM on group contributions by comparing the two levels of inequality, both with and without the MAM. In Figure 1, we observe that in the baseline treatments, the average group contribution decreases after the first two rounds and continues to decline until the end of the game (Mann-Kendall trend test, $p < 0.001$). Conversely, average contributions in the MAM treatments consistently increase over rounds

[4]As in the original paper, we observe a clear convergence to the social optimum. The means and standard deviations were: without the MAM, $G = 4.17$ $sd = 5.07$ and, with the MAM, $G = 26.88$ $sd = 13.94$ and a last time period with $G = 40$ the social optimum.

(Mann-Kendall trend test, $p < 0.001$). These opposite tendencies both suggest a learning effect. In the first case, participants learn to avoid being exploited by reducing their contributions, while in the second case, they learn to effectively use the mechanism. Consequently, contributions with and without the MAM are significantly different in both inequality conditions (Wilcoxon rank-sum test, $p < 0.001$).

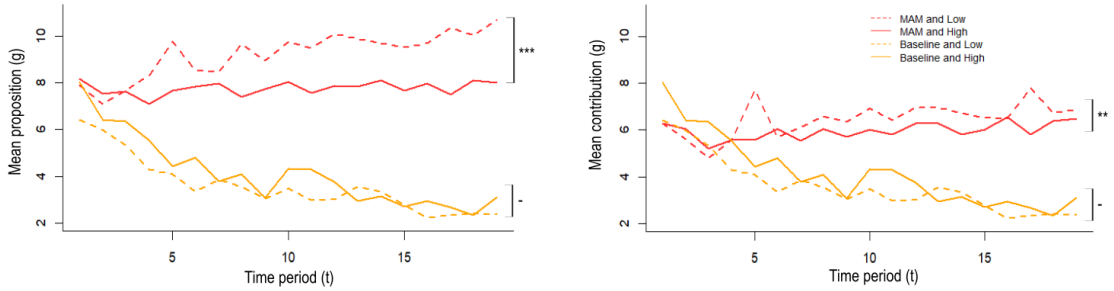
Figure 1: Mean group contributions over time per treatment



Note: The figure displays Wilcoxon rank-sum test results, with the following p-value significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Result 1. *The MAM significantly increases contributions to the public good, although the increase is less than what is predicted by BEWDS.*

Figure 2: Mean individual proposals (left) and contributions (right) over time per treatment and inequality level



Note: The figure displays Wilcoxon rank-sum test results, with the following p-value significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Orange lines in both graphs display contributions without the MAM.

Contributions in the two inequality treatments do not differ significantly without the MAM: the average contribution is $g = 3.59$ under low inequality and $g = 4.41$ under high inequality (Wilcoxon rank-sum, $p = 0.211$) as predicted by the theory. However, with the

MAM, there is a significant difference: the average contribution is $g = 6.49$ under low inequality and $g = 5.98$ under high inequality (Wilcoxon rank-sum, $p < 0.001$). While this difference is lower than in the theory which predicted that the contributions should equal the wealth of the poorest player, the tendency is confirmed and players contribute less when the inequalities are high.

Comparing these results with those of our replication without inequality and those of the original study of Masuda et al. (2014), it is evident that simply introducing inequality significantly impacts the effectiveness of the mechanism, more so than the magnitude of the inequality itself. Theoretically, introducing low inequality should reduce the contribution from 40 tokens (100% of endowment) to 36 tokens (90% of endowment) and finally to 20 tokens (50% of endowment) with high inequality. However, in the experiment, the contributions in the last period in our replication and in Masuda et al. (2014) were indeed 40 tokens, compared to only 13.81 tokens (35%) for low inequality and 13.00 tokens (33%) for high inequality.

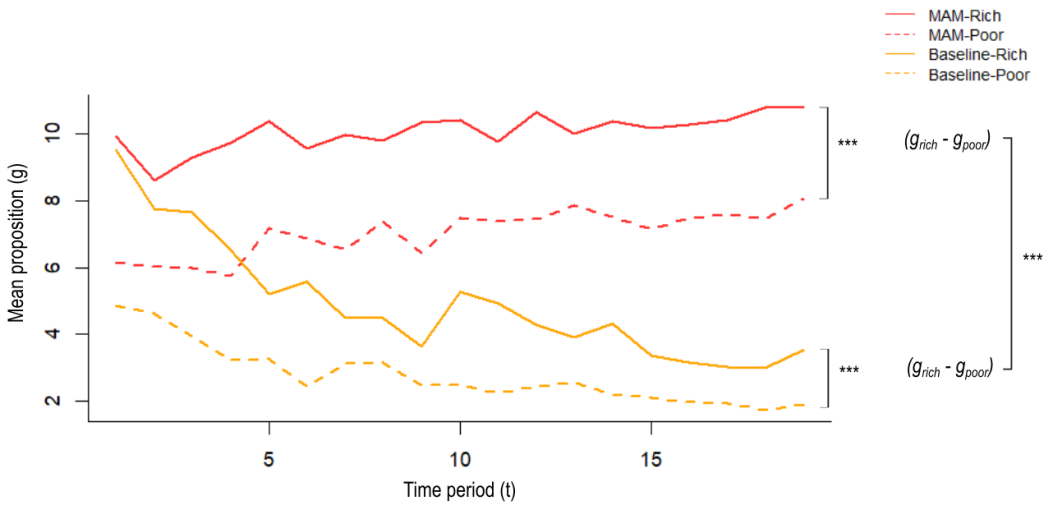
Result 2. *Endowment inequalities strongly weaken the effectiveness of the MAM.*

Result 3. *The extent of endowment inequality has minimal influence on contributions.*

However, upon examining the proposals in Figure 2, it becomes evident that the MAM significantly increases the amounts individuals propose to contribute in the first stage (Wilcoxon rank-sum, $p < 0.001$) across both inequality conditions, serving as a safety net for players inclined to cooperate. Specifically, proposals are notably higher in the low inequality treatment (Wilcoxon rank-sum, $p = 0.006$). In the low inequality condition, proposals reach 51.4% of the equilibrium (9.26 units with the MAM compared to 3.59 without), while in the high inequality treatment, they reach 77.8% (7.78 units with the MAM compared to 4.63 without). This underscores that even introducing minimal inequalities has a profound impact on behaviors and on the effectiveness of the MAM.

Without the MAM, the rich player proposes higher contributions, specifically $g^{rich,low} = 3.99$ and $g^{rich,high} = 5.98$, compared to the poor player, with $g^{poor,low} = 3.20$ and $g^{poor,high} = 2.31$ (Wilcoxon rank-sum, $p < 0.001$). This disparity increases further with the MAM, where the mean contributions for the rich are $g^{rich,low} = 10.07$ and $g^{rich,high} = 10.05$, and for the poor are $g^{poor,low} = 8.44$ and $g^{poor,high} = 5.52$. These

Figure 3: Mean individual proposals (contributions without MAM) over time per treatment and role



Note: The figure displays Wilcoxon rank-sum test results, with the following p-value significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

proposed contributions are lower than the predicted proposals under the MAM ($g_i^* = w_1$), indicating that the effective efficiency of the MAM is lower than anticipated. On average, under conditions of low inequality, the average proposal is slightly above 50% of the predicted value, whereas under high inequality, it rises to 77.85% of the predicted proposal. Note, however, that under high inequality, the average proposal of the rich closely aligns with the predicted value. Finally, we observe that, regardless of the presence or absence of the MAM, individuals with higher endowments tend to propose higher contributions, contrary to the initial prediction. We delve into the motivations in Section 4.2.

Result 4. *The rich players propose more with and without the MAM than the poor players. This difference increases with the MAM.*

It implies that the increase in inequalities has an opposite effect on rich and poor players reducing poor players' proposals and increasing rich players' proposals.

4.2 Inequality aversion predictions

In this subsection we test the prediction of the Fehr and Schmidt (1999) model, that the strength of the guilt aversion of the rich player, β_2, increases its contribution. Additionally, we expect this effect to influence not only the final contribution but also the proposals and voting decisions. To analyze this prediction, we first compute the α_i and β_i parameters of each player as described in Section 3.2. Figure 4 displays the joint distribution of α

and β through our sample. We observe a strong heterogeneity in the inequality aversion parameters. To compute the effect of the inequality aversion, we first control for the compliance with the Fehr and Schmidt (1999)'s rationality conditions: $\beta < 1$ and $\beta \leq \alpha$. These conditions exclude advantaged players who are willing to pay more than 1 unit to reduce inequalities by giving 1 unit to the disadvantaged player, and players who would suffer more from advantageous than disadvantageous inequalities. These two scenarios are described by Fehr and Schmidt (1999) as highly implausible. Applying these conditions reduces our dataset from 310 to 180, with 35 subjects violating $\beta < 1$, 76 subjects violating $\beta \leq \alpha$ and 19 violating both. Accordingly, we will conduct each analysis with and without these conditions. However, it is notable that, similar to Blanco et al. (2011), where at least 38% of the subjects violated at least one condition, a significant portion of our data (34% of the subjects) does not support the previous two assumptions.

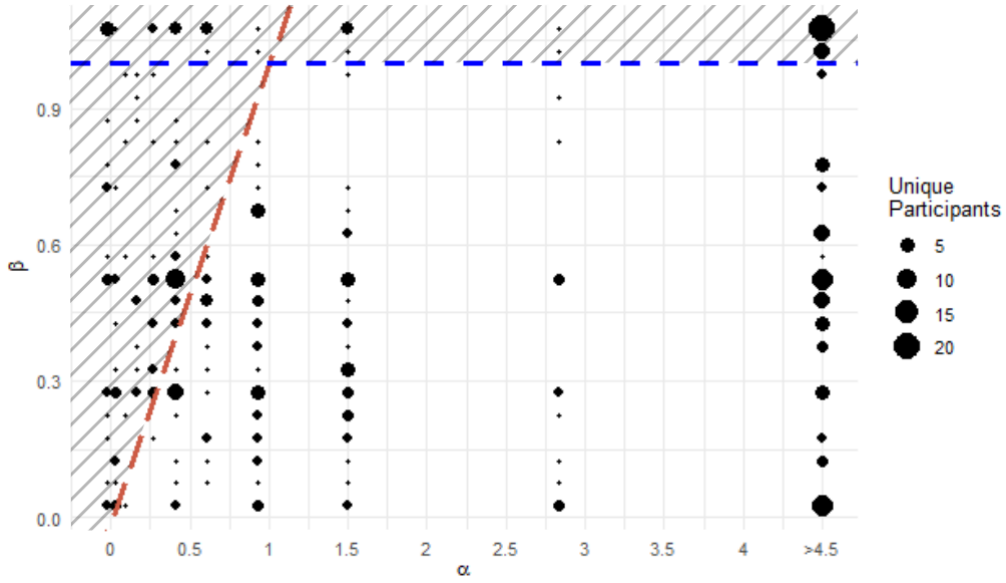


Figure 4: Distribution of individuals per α and β.

Note: As in Blanco et al. (2011) we display $\alpha > 4.5$ for more readability but we use the true values for the tests. The line $\beta = 1$ corresponds to the first Fehr and Schmidt (1999)'s condition. The line $\alpha = \beta$ corresponds to a similar inequality aversion for advantageous and disadvantageous situations, and to the second condition. The dashed part are the participants excluded when the two Fehr and Schmidt (1999)'s conditions are satisfied. Appendix A.3 shows that we find different distribution to both Fehr and Schmidt (1999) and Blanco et al. (2011) which themselves already observed different distributions.

Before examining the effect of β_2, we checked the distribution across treatments and found no significant differences in the distribution of β between the different samples (Kolmogorov-Smirnov tests: MAM vs. baseline, p = 0.08; Low vs. High, p = 0.85). Additionally, we tested the interaction of Treatment and Inequality level and found no

significant difference in the means of beta across the different groups (Kruskal-Wallis test: p = 0.3187). Thus, we can assume that running the DG and UG after the main task did not affect the parameter estimations of Fehr and Schmidt (1999)'s β.

To control for the effect of β_2 on the proposal behaviors of the rich players, we rely on the following linear regression with aggregated data at the individual level[5]:

$$g_i = \lambda_0 + \lambda_1.\beta_i + \lambda_3.\text{ineq\_level} + \lambda_4.\text{MAM} + \epsilon_i$$

where g_i as the individual mean proposal, β_i as the advantageous inequality aversion, *ineq_level* as the inequality level, and *MAM* as the presence of the mechanism. We initially ran the regression with only β and then iteratively added the control variables. Additionally, we conducted all regressions using both the entire dataset (see Appendix A.4.1) and a subset of subjects adhering to Fehr and Schmidt (1999)'s rationality conditions (see Appendix A.4.2).

Overall, there is no significant effect of β_i on the proposals of the rich players. However, if we consider only the subset of participants being consistent with the Fehr and Schmidt (1999)'s rationality conditions (66% of the rich players), we observe a strong positive effect of β_2 on proposals (+5.7 units proposed, $p < 0.05$).

Result 5. *Only for the players who satisfy the Fehr and Schmidt (1999)'s conditions, the guilt aversion strongly increase the proposals with or without the MAM.*

Conclusions about the predictability of the guilt parameter β_2 of the rich player's contributions are mixed, as they are not generalizable to all our subjects. As part of our effort to falsify Propositions 4 and 5, which define rich players' equilibria based on their low or high guilt types using the threshold $\beta > 1 - \lambda$, we demonstrate that players' decisions differ significantly from these predictions (Wilcoxon signed-rank test, $p < 0.001$). Additionally, we show by replacing β in the regression of the Appendix A.4.1, that the type has no influence on the Stage 1 decisions of the rich players (OLS with fixed effect at the session level, $p = 0.46$).

For comprehensiveness, we also test the effect of α on the rich players and the effect of both parameters on the poor players (see Appendix A.4.1 and A.4.2 for the rich players and A.5 for the poor players). As expected, disadvantageous inequality aversion does not influence

[5]A panel regression with round fixed effects clustered at the session level led to the same results.

the proposals or contributions of the rich players. Similarly, at the sample level, neither advantageous nor disadvantageous inequality aversion has a significant effect on the poor players[6]. However, when focusing on poor players who satisfy the conditions outlined by Fehr and Schmidt (1999) (50% of them), we find that, contrary to the theoretical framework, their alpha increases significantly their proposals (+0.254 units, $p < 0.01$) but does not affect their contributions. This suggests that poor players believe proposing more will lead to increased collective contributions. If they anticipate that rich players will propose more and subsequently disapprove, it effectively boosts the collective contribution.

Overall, our analysis reveals a nuanced relationship between inequality aversion and the observed behaviors. While some elements of the theoretical predictions are supported, such as the impact of guilt aversion on rich players' proposals and envy aversion on poor players' proposals under some assumptions, it appears that these conclusions are not generalized to our data. Given these mixed results, it becomes essential to further examine how individuals make approval decisions under the MAM to understand the observed behaviors.

4.3 Approval decisions

In Figure 2, we observe a significant disparity between proposed and realized contributions, highlighting the crucial role of approval decisions. In this subsection, we examine approval decisions in relation to theoretical predictions. The lemma 1 suggests that min-proposers (lowest contributors) systematically approve while the max-proposers (highest contributors) always disapprove. We find that min-proposers tend to behave rationally, i.e. maximizing their profits. Conversely, max-players often approve, thereby foregoing profit maximization.

Result 6. *Under the MAM, 96% of the min-contributors approve, confirming the first part of Lemma 1. However, only 66.7% of the max-contributors disapprove, in contradiction with the second part of Lemma 1.*

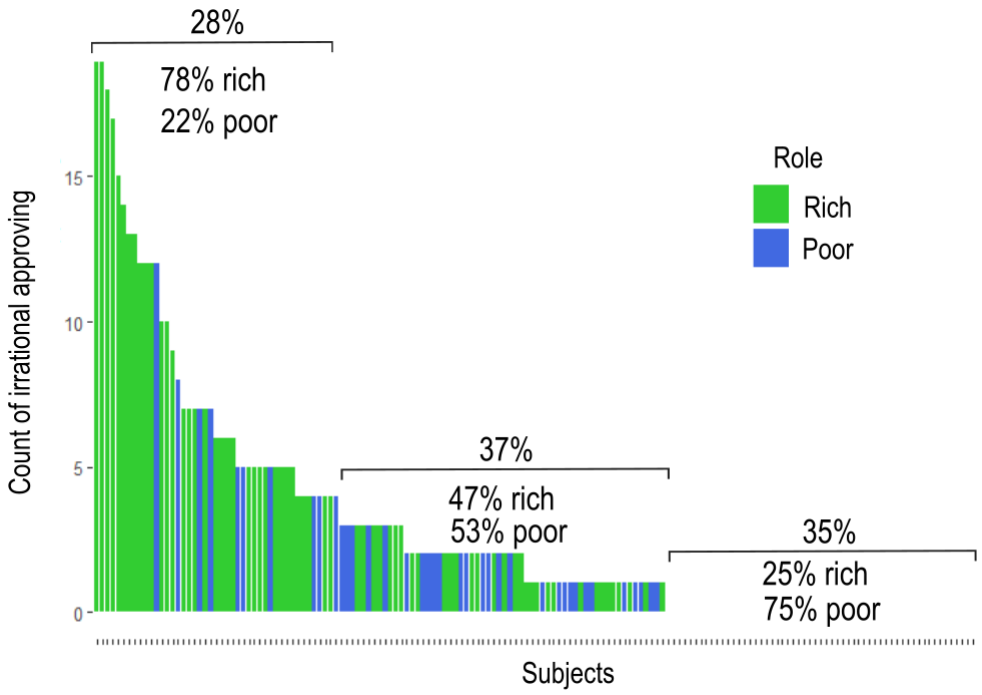
To gain clearer insights into underlying behavior, let us break it down by rich and poor players. We observe that in 61.4% of instances, the rich player proposes more than the poor player, whereas in 30.7% of cases, it is the poor player who proposes more. Now, examining how players approve in disadvantageous situations based on their role:

[6]The significance of the α parameter for the poor players drops when we add the control variables.

- When a rich player proposes more, they approve in 38.3% of these cases, accounting for 23.5% of all decisions.

- Conversely, when the poor player proposes more, they approve in 23.2% of these cases, which represents only 7.12% of all decisions.

Figure 5: Count of irrational approvals by participant



Note: Each dash (n=164) on the x-axis corresponds to one player of the MAM treatments. The vertical bars correspond to the number of time a given player decided to approve a disadvantageous vector. The numbers displayed the percent of players overall and per role displayed on the x-axis.

The apparent departure from profit maximization behavior is primarily driven by rich players, as illustrated in Figure 5. In Table 2, we employ a Logistic regression with Robust Cluster standard errors at the individual level[7] to analyze the effect of role, difference in proposals, NLE score, parameters α[8], and β. Additionally, we examine the interaction between β and the "rich" role, along with the round number to account for learning effects. This regression focuses solely on players who encountered at least one instance where they proposed more than the other player.

Result 7. *Rich players approve disadvantageous proposal vectors significantly more fre-*

[7]Due to uneven representation of individuals in the dataset (only rounds where they faced a disadvantageous vector are included), we control for robust cluster standard errors at the participant level. This method ensures the robustness of our results. Significant differences in contribution ($p < 0.1$) and the β variable ($p < 0.1$) diminished, suggesting they were influenced by over-represented individuals.

[8]We included α in this regression because this situation involves deciding in a potentially disadvantageous scenario, even though the outcome could eventually be higher than the other player's outcome.

Table 2: Logistic Regression Results with Robust Cluster Standard Error

	(1)		(2)		(3)	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	-0.55*	0.24	-0.23	0.34	1.64	1.60
rolerich	0.79**	0.29	0.91**	0.29	1.33*	0.56
diff_of_contrib			-0.05	0.04	-0.05	0.04
NLE_score					-1.01	0.59
alpha					-0.00	0.04
beta					1.23	0.68
round_number	-0.07***	0.01	-0.08***	0.01	-0.08***	0.01
rolerich*beta					-0.15	0.83
Log Likelihood	-861.98		-851.99		-823.87	
Observations	1417		1417		1417	
AIC	1729.96		1711.99		1663.74	

Note: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. Model (1) basically test the effect of the role with the cluster correction with the learning fixed effect. Model (2) introduces the difference of contribution of the players. Model (3) introduces the individual characteristics.

quently than poor players.

Consistent with the results on proposals and contributions, at the sample level, we find that advantageous inequality aversion (β) does not predict the approval decision of disadvantageous vectors. However, being the richer player increases the probability of approving a disadvantageous vector. Even when considering only the subset of rich players satisfying Fehr and Schmidt (1999)'s rationality conditions, we find no effect of β_2 on the rate of irrational disapproval. The variable 'round number' negatively impacts the probability of approval, suggesting a learning effect in which players increasingly prioritize profit maximization as they gain experience in the game. Furthermore, the NLE score and the difference in proposals between the players do not affect the probability of approving.

Result 8. *When facing a disadvantageous proposal vector, 38.3% of the rich players and 23.2% of the poor players choose to approve.*

A plausible interpretation is that while rich players are willing to contribute more than poor players, they conscientiously weigh the benefits and costs of doing so. This suggests they are cautious about exerting higher relative effort compared to poor players.

4.4 Additional (exploratory) results

The inequality aversion theory of Fehr and Schmidt (1999) fails to describe the approval decisions, i.e. the fact that rich players often approve contribution pairs in which they proposed the largest contribution. In this subsection we explore an alternative explanation[9] based on beliefs regarding how the rich players should contribute compared to the poor players.

Table 3: Comparison of proposals/contributions to the beliefs about partner's proposals/contributions

	(1) Rich			**(2) Poor**		
Ineq.	Contr/Prop.	Belief RvP	p	Contr/Prop.	Belief PvR	p
High-Baseline	5.98 (7.67)	2.35 (3.16)	***	2.31 (3.20)	5.62 (9.21)	***
Low-Baseline	3.99 (5.72)	3.49 (5.12)	**	3.20 (4.87)	5.30 (7.05)	***
High-AM	10.05 (6.12)	4.89 (3.87)	***	5.52 (3.55)	9.02 (10.35)	***
Low-AM	10.07 (6.50)	8.51 (6.31)	***	8.44 (5.83)	10.40 (7.68)	***

Note: Standard deviations are shown in parentheses next to the means. Significance levels: ** $p < 0.01$, *** $p < 0.001$. Columns (1) display the first stage decision of rich players and if it is significantly higher than their beliefs about the poor players' first stage decisions. Columns (2) display the first stage decision of poor players and if it is significantly lower than their beliefs about the rich players' first stage decisions. We use a Wilcoxon test to compare the contributions to the beliefs about the partner.

To assess the beliefs of the rich players about themselves and confirm that beliefs are a motivation in this setting, Table 3 establishes the following comparisons:

(1) Proposals/contributions of rich players compared to their beliefs about poor players' proposals/contributions. Rich contribute significantly more than what they expect from the poor (Wilcoxon signed-rank, $p < 0.001$), suggesting that they share the belief that rich should contribute more than poor.

(2) Whether poor players contribute less than what they expect rich players to contribute. This comparison is also consistent (Wilcoxon signed-rank, $p < 0.001$), indicating they align with the belief that rich players should contribute more than poor players. Finally, we find that the beliefs of rich players about poor players are significantly lower than the beliefs of poor players about rich players (Wilcoxon signed-rank, $p < 0.001$).

Result 9. *Both rich and poor share the belief that the rich player should contribute more than the poor player by contributing (proposing) accordingly.*

[9]This exploratory research has not been preregistered but aims to enhance understanding of the observed behaviors. Further research would therefore be needed to confirm the results of this exploratory analysis.

One possible explanation for the effect of beliefs on behaviors while inequality aversion has no effect is that considering, as rich or poor player, that the rich player has to contribute more is not for the purpose of reducing inequalities but only because it is perceived as fair that the efforts are correlated with the wealth. Martinangeli (2021) recently suggested, in a similar setting involving a rich and a poor contributor, that both first-order and second-order beliefs about the contributions of the roles influence decisions regarding unequal contributions. Both players generally expect the rich player to contribute more than the poor player.

5 Summary and Concluding Remarks

In this experimental study, we investigated the impact of the Minimum Approval Mechanism (MAM) on public good provision, focusing on the effects of endowment and preferences heterogeneity. By conducting laboratory experiments, we analyzed how income inequality influences contributions and decision-making within the framework of the MAM. Our investigation into the Minimum Approval Mechanism confirms its performance in Pareto-improving cooperation for public good provision. Although the mechanism was efficient in an environment with homogeneous individuals, introducing even a very low endowment inequality is sufficient to strongly reduce cooperation. This reduced effectiveness does not decrease much further with higher inequalities. The lower performance of the MAM with endowment-inequalities, is primarily attributed to lower-than-expected proposals from participants, although the MAM provides a safety net for cooperative behavior. We relied on inequality aversion to explain differences in behaviors between rich and poor players and found that, as predicted by the theoretical framework, only the advantageous inequality aversion of the rich players has a positive effect on their contributions. Nevertheless, this effect is observable only for the subset of individuals (65% of the sample) whose decisions in the dictator and ultimatum game satisfy the rationality conditions of the fairness theory of Fehr and Schmidt (1999).

We found that the irrational approval decisions, i.e. approving disadvantageous proposals, were a key difference in effectiveness between the theoretical predictions and the experimental results. A large part of rich players decided to frequently accept to contribute more than the poor players. Surprisingly, inequality aversion is not the motivation behind

this behavior. The data provides strong indications that participants' stated beliefs about role expectations significantly influence their behavior in this asymmetric game setting. Rich and poor players share the belief that rich players should contribute more. It appears that this 'fairness' principle, rather than a drive to mitigate inequality, motivates players to differentiate their contributions based on their roles, aiming not for equality but for increasing the provision of the public good. This finding aligns rather with the guilt aversion hypothesis described by Dufwenberg and Kirchsteiger (2004); Battigalli and Dufwenberg (2007, 2009), which differs from the guilt described by Fehr and Schmidt (1999). In this case, guilt aversion refers to the disutility of letting down another's first-order beliefs about what one should do. The rich player could feel guilty if she thinks they both share the belief about her relative contribution. Another potential explanation is that individuals seem guided by a distributional norm, believing that the rich should contribute more than the poor. Accordingly, a rich player contributes more without expecting poor players to contribute as much. This behavior is mirrored by poor players, who contribute less than they expect the rich to contribute. This implies that individuals consider what they would do in the other role and act accordingly, following a categorical imperative (Kant, 1785) and suggesting that this rule is universalizable. Thus, individuals contribute following a Kantian norm (Alger and Weibull, 2016; van Leeuwen and Alger, 2023). The behaviors we observe also align with the findings of López-Pérez and Vorsatz (2010), who demonstrated that individuals prefer to comply with social norms to avoid negative feedback and disapproval from others.

The nuanced behavioral responses to heterogeneity underscore the importance of considering initial inequalities and participant beliefs in designing and implementing mechanisms for mitigating social dilemma. Moreover, the findings from our study reinforce the relevance of the MAM and identify one weakness which has to be studied by considering reinforcing informations about the safety for cooperators induced by the mechanism. This study brings insights for future research, particularly in understanding the underpinnings of cooperative behavior in the presence of endowment inequalities. The limited predictive power of inequality aversion models in our findings suggests that alternative frameworks, which incorporate beliefs and motivations more explicitly, might offer deeper insights, especially regarding the relevance of moral preferences in this context. Ultimately, advancing our understanding in these areas will enhance the design of more effective, equitable, and

context-sensitive mechanisms for public good provision.

Originally, one of the motivations for setting the minimum contribution benchmark in case of disapproval was the willingness to build a mechanism leading to a Pareto-improvement without forcing any individuals to contribute more than what they proposed. However, the insights on shared beliefs and the Kantian norm of role-specific contributions suggest that rich individuals are motivated to contribute more than poor individuals. This implies that individuals could be better off when there is an asymmetry of contribution relative to the initial endowment. This can be understood as a compensatory contribution inequality, triggered by a sense of guilt, to offset the inequality of endowment. Extending the MAM to a proportional framework, where players contribute the lowest proposed proportion of their endowment in the case of disapproval, could provide valuable insights and address the mechanism's weaknesses.

Instruments like the MAM are particularly applicable in the context of climate change, where countries face unequal endowments or historical emissions but must agree on minimum contribution levels for each party. These contributions could be defined in relative terms rather than fixed absolute amounts, as explored in this paper. This perspective also opens potential avenues to refine the mechanism, enhancing its practicality and feasibility for real-world implementation.

6 Conflict of Interest and AI-use Declarations

The authors declare no conflict of interest regarding this work. The authors used generative AI tools to improve the writing and readability of the manuscript. All content was reviewed and edited by the authors, who take full responsibility for the final version of this paper.

7 Acknowledgements

We thank the participants of the Game Theory and Experiments 2024 Workshop at the University of Waseda, Japan, and the Association of French Experimental Economics 2024 conference for their valuable discussions. Special thanks to Brice Magdalou and Mickaël Beaud for their comments on the theoretical framework, and to Dimitri Dubois and Sarah Feti for their assistance with lab management.

8 Supplementary data

The dataset, the R code for the data analysis and the oTree code for replication will be made available on Github as online supplementary materials.

References

Alger, I. and J. W. Weibull (2016). Evolution and Kantian morality. *Games and Economic Behavior 98*, 56–67.

Battigalli, P. and M. Dufwenberg (2007). Guilt in Games. *American Economic Review 97*(2), 170–176.

Battigalli, P. and M. Dufwenberg (2009). Dynamic psychological games. *Journal of Economic Theory 144*(1), 1–35.

Blanco, M., D. Engelmann, and H. T. Normann (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior 72*(2), 321–338.

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

Chen, D. L., M. Schonger, and C. Wickens (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance 9*, 88–97.

Chen, Y. and C. R. Plott (1996). The Groves-Ledyard mechanism: An experimental study of institutional design. *Journal of Public Economics 59*(3), 335–364.

Dufwenberg, M. and G. Kirchsteiger (2004). A theory of sequential reciprocity. *Games and Economic Behavior 47*(2), 268–298.

Falkinger, J. (1996). Efficient private provision of public goods by rewarding deviations from average. *Journal of Public Economics 62*(3), 413–422.

Falkinger, J., E. Fehr, S. Gächter, and R. Winter-Ember (2000). A Simple Mechanism for the Efficient Provision of Public Goods: Experimental Evidence. *American Economic Review 90*(1), 247–264.

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

Fehr, E. and K. Schmidt (1999). A Theory of Fairness, Competition and Cooperation. *Quarterly Journal of Economics 114*, 817–868.

Groves, T. (1973). Incentives in Teams. *Econometrica 41*(4), 617–631. Publisher: [Wiley, Econometric Society].

Groves, T. and J. Ledyard (1977). Optimal Allocation of Public Goods: A Solution to the "Free Rider" Problem. *Econometrica 4*, 783–809.

Halberda, J. and D. Odic (2015). Chapter 12 - The Precision and Internal Confidence of Our Approximate Number Thoughts. In D. C. Geary, D. B. Berch, and K. M. Koepke (Eds.), *Mathematical Cognition and Learning*, Volume 1 of *Evolutionary Origins and Early Development of Number Processing*, pp. 305–333. Elsevier.

Kalai, E. (1981). Preplay negotiations and the prisoner's dilemma. *Mathematical Social Sciences 1*(4), 375–379.

Kant, I. (1785). *Groundwork for the Metaphysics of Morals.* New Haven, CT: Yale University Press.

Laussel, D. and T. R. Palfrey (2003). Efficient Equilibria in the Voluntary Contributions Mechanism with Private Information. *Journal of Public Economic Theory 5*(3), 449–478. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9779.00143.

Ledyard, J. (1995). Public Goods: A Survey of Experimental Research. In *Handbook of Experimental Economics* (Princeton University Press, Princeton, NJ, 2016 ed.), Volume 2.

López-Pérez, R. and M. Vorsatz (2010). On approval and disapproval: Theory and experiments. *Journal of Economic Psychology 31*(4), 527–541.

Martinangeli, A. F. M. (2021). Do what (you think) the rich will do: Inequality and belief heterogeneity in public good provision. *Journal of Economic Psychology 83*, 102364.

Masuda, T., Y. Okano, and T. Saijo (2014). The Minimum Approval Mechanism Implements the Efficient Public Good Allocation Theoretically and Experimentally. *Games and Economic Behavior*.

McKelvey, R. D. and T. R. Palfrey (1998). Quantal Response Equilibria for Extensive Form Games. *Experimental Economics 1*(1), 9–41.

Nagel, R. (1995). Unraveling in Guessing Games: An Experimental Study. *American Economic Review 85*(5), 1313–1326. Publisher: American Economic Association.

Renou, L. and K. H. Schlag (2011). Implementation in minimax regret equilibrium. *Games and Economic Behavior 71*(2), 527–533.

Saijo, T., Y. Okano, and T. Yamakawa (2011). The Approval Mechanism Experiment: A Solution to Prisoner's Dilemma. *undefined*.

Siegler, R. S. and J. E. Opfer (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychological Science 14*(3), 237–243.

van Leeuwen, B. and I. Alger (2023). Estimating Social Preferences and Kantian Morality in Strategic Interactions.

Varian, H. R. (1994). A Solution to the Problem of Externalities When Agents Are Well-Informed. *American Economic Review 84*(5), 1278–1293. Publisher: American Economic Association.

Yao, K. S. W., E. Lavaine, and M. Willinger (2022). Does the approval mechanism induce the efficient extraction in common pool resource games? *Social Choice and Welfare 58*(1), 111–139.

List of Figures

List of Tables

A Appendix

A.1 Proofs of lemmas

A.1.1 Proof of Lemma 1

Proof. Comparing the approval payoff $\pi_i(\underline{g}, \overline{g})$ with the disapproval payoff $\pi_i^D(\underline{g}, \underline{g})$ for player i, who proposes the lowest contribution, we find that player i approves if $\pi_i - \pi_i^D \geq 0$:

$$\pi_i(\underline{g}, \overline{g}) - \pi_i^D(\underline{g}, \underline{g}) \geq 0 \iff w_i - \underline{g} + \gamma(\underline{g} + \overline{g}) - w_i + \underline{g} - 2\gamma\underline{g} \geq 0$$
$$\iff \underline{g} \leq \overline{g}$$

This is true by definition, as $\underline{g}$ is the minimum contribution. □

A.1.2 Proof of Lemma 2

Proof. Given that the conditions of the previous propositions are met, asymmetric proposition vectors do not persist to the second stage. Conversely, by the construction of the MAM, the profits $\pi_i(g, g) = \pi_i^D(g, g)$ are equals, implying symmetric proposition vectors survive the second stage. □

A.1.3 Proof of Lemma 3

Proof. Players' approval decisions depend on comparing their payoffs under approval versus disapproval. A single player's disapproval is enough to ensure that the Minimum proposal is implemented.

(i) Player 2 never approves any scenario where $g_2 > w_2 - (w_1 - g_1)$, as this would reverse the inequality and activate her aversion to disadvantageous inequality, α_2.

(ii) For $g_2 \leq w_2 - (w_1 - g_1)$ and $g_1 < g_2$, Player 1's approval is guaranteed if:

$$U_1(g_1, g_2) - U_1^D(g_1, g_1) \geq 0 \iff (g_2 - g_1)(\gamma + \alpha_1) \geq 0, \tag{5}$$

which is always true. Player 2's decision to approve is contingent upon her level of guilt aversion, β_2, such that she approves if $\beta_2 > 1 - \gamma$.

(iii) For $g_1 > g_2$, Player 1 disapproves because:

$$U_1(g_1, g_2) - U_1^D(g_2, g_2) < 0 \iff (g_1 - g_2)(1 + \alpha_1 - \gamma) > 0, \tag{6}$$

which holds by the assumption regardless of positive aversion to disadvantageous inequality, α_1. □

A.2 F&S predictions without endowment heterogeneity

Proposition 6. *Under the MAM with BEWDS, the equilibrium contribution pair is the social optimum (w, w) with endowment equality, regardless of the inequality aversion heterogeneous preferences.*

Proof. In the scenario where $g_1 = g_2$, players are indifferent between approving and disapproving since it leads to the same outcome. Due to this equality, their utility functions are $U_i(g, g) = w - g + 2\gamma g$, neutralizing the α_i and β_i parameters due to the payoff equality $\pi_1 = \pi_2$.

For $g_1 > g_2$, player 1 disapproves if $U_1^D(g_2, g_2) = w - g_2 + 2\gamma g_2 > U_1^A(g_1, g_2) = w - g_1 + \gamma(g_1 + g_2) - \alpha_1(g_1 - g_2)$, which is systematically true for any $\alpha_1 \geq 0$ and $g_1 > g_2$. Thus, any asymmetric proposals lead the player contributing g_1 to disapprove, driving the outcome towards symmetric contributions.

Accordingly, in the normal form game, stage 1, both players will contribute the same amount g. We can write their utility functions as: $U_1(g, g) = w_1 - g + 2\gamma g - \alpha_1(w_2 - w_1)$ and $U_2(g, g) = w_2 - g + 2\gamma g - \beta_2(w_2 - w_1)$, both increasing in g when $\gamma > \frac{1}{2}$, which is true by assumption. Due to the echelon structure, it converges to the social optimum (w, w) with $g = w$. □

A.3 F&S Parameters distribution

Table 4: Comparison of Distributions Across Studies

Interval	F&S (%)	Blanco et al. (%)	Our Data (%)
Alpha Distribution			
$\alpha < 0.4$	30	31	21
$0.4 \leq \alpha < 0.92$	30	33	20
$0.92 \leq \alpha < 4.5$	30	23	29
$4.5 \leq \alpha$	10	13	31
Beta Distribution			
$\beta < 0.235$	30	29	21
$0.235 \leq \beta < 0.5$	30	15	28
$0.5 \leq \beta$	40	56	51

Note: Using χ^2 tests, we find that our data are only different from both other samples in α with $p < 0.01$ and not in β. The distribution of β of the two other samples are significantly different with $p < 0.05$.

A.4 Rich's α and β effects

A.4.1 With all the subjects

Table 5: Regression Results for mean_prop: Rich players

	(1)	(2)
alpha	0.027 (0.101)	0.057 (0.101)
beta	0.405 (1.298)	0.500 (1.295)
ineq_levellow	-0.962 (0.779)	-0.891 (0.783)
trtbaseline	-5.105*** (0.779)	-5.118*** (0.773)
genderMasculin		1.264 (0.788)
NLE_score		1.729 (1.408)
Constant	10.269*** (0.915)	4.799 (3.950)
Observations	155	155
R2	0.231	0.255
Adjusted R2	0.211	0.224
Residual Std. Error	4.836 (df = 150)	4.794 (df = 148)
F Statistic	11.279*** (df = 4; 150)	8.421*** (df = 6; 148)

Note: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Table 6: Regression Results for mean_g_i: Rich players

	(1)	(2)
alpha	0.103 (0.090)	0.117 (0.091)
beta	0.931 (1.150)	0.973 (1.160)
ineq_levellow	-1.309 (0.691)	-1.271 (0.702)
trtbaseline	-1.881** (0.690)	-1.887** (0.692)
genderMasculin		0.598 (0.706)
NLE$_score$		0.880 (1.261)
Constant	6.752*** (0.810)	3.994 (3.537)
Observations	155	155
R2	0.083	0.092
Adjusted R2	0.059	0.055
Residual Std. Error	4.285 (df = 150)	4.294 (df = 148)
F Statistic	3.411* (df = 4; 150)	2.492* (df = 6; 148)

Note: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

A.4.2 With Fehr and Schmidt (1999)'s conditions

Table 7: Regression Results for mean_prop on FS subset: Rich players

	(1)	(2)
alpha	0.083 (0.120)	0.111 (0.122)
beta	4.833* (2.196)	5.742* (2.265)
ineq_levellow	-0.287 (0.929)	-0.178 (0.936)
trtbaseline	-4.556*** (0.947)	-4.409*** (0.954)
genderMasculin		1.497 (0.967)
NLE_score		1.007 (1.639)
Constant	7.986*** (1.127)	3.954 (4.798)
Observations	102	102
R2	0.219	0.241
Adjusted R2	0.186	0.193
Residual Std. Error	4.666 (df = 97)	4.646 (df = 95)
F Statistic	6.782*** (df = 4; 97)	5.035*** (df = 6; 95)

Note: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Table 8: Regression Results for mean_g_i on FS subset: Rich players

	(1)	(2)
alpha	0.148 (0.107)	0.175 (0.108)
beta	5.632** (1.948)	6.481** (2.005)
ineq_levellow	-0.680 (0.824)	-0.574 (0.829)
trtbaseline	-1.515 (0.840)	-1.374 (0.844)
genderMasculin		1.398 (0.857)
NLE_score		0.976 (1.451)
Constant	4.597*** (0.999)	0.729 (4.248)
Observations	102	102
R2	0.134	0.162
Adjusted R2	0.098	0.109
Residual Std. Error	4.139 (df = 97)	4.113 (df = 95)
F Statistic	3.739** (df = 4; 97)	3.058** (df = 6; 95)

Note: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

A.5 Poor's α and β effects

A.5.1 With all the subjects

Table 9: Regression Results for mean_prop: Poor players

	(1)	(2)
alpha	0.116* (0.059)	0.114 (0.059)
beta	0.839 (0.694)	0.852 (0.693)
ineq_levellow	1.924*** (0.479)	2.012*** (0.481)
trtbaseline	-3.940*** (0.499)	-4.009*** (0.502)
genderMasculin		-0.138 (0.481)
NLE_score		1.414 (0.917)
Constant	4.970*** (0.650)	1.219 (2.571)
Observations	155	155
R2	0.404	0.414
Adjusted R2	0.388	0.390
Residual Std. Error	2.975 (df = 150)	2.970 (df = 148)
F Statistic	25.411*** (df = 4; 150)	17.398*** (df = 6; 148)

Note: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Table 10: Regression Results for mean_g_i: Poor players

	(1)	(2)
alpha	0.059 (0.051)	0.057 (0.051)
beta	0.777 (0.601)	0.785 (0.601)
ineq_levellow	1.245** (0.414)	1.301** (0.418)
trtbaseline	-2.624*** (0.431)	-2.653*** (0.435)
genderMasculin		-0.283 (0.417)
NLE_score		0.866 (0.796)
Constant	4.200*** (0.562)	2.004 (2.230)
Observations	155	155
R2	0.283	0.291
Adjusted R2	0.264	0.262
Residual Std. Error	2.573 (df = 150)	2.576 (df = 148)
F Statistic	14.822*** (df = 4; 150)	10.127*** (df = 6; 148)

Note: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

A.5.2 With Fehr and Schmidt (1999)'s conditions

Table 11: Regression Results for mean_prop on FS subset: Poor players

	(1)	(2)
alpha	0.236** (0.083)	0.254** (0.086)
beta	2.535 (1.348)	2.295 (1.369)
ineq_levellow	1.803* (0.688)	1.768* (0.694)
trtbaseline	-4.251*** (0.707)	-4.173*** (0.713)
genderMasculin		0.002 (0.699)
NLE_score		1.784 (1.514)
Constant	3.939*** (0.892)	-0.947 (4.258)
Observations	78	78
R2	0.489	0.498
Adjusted R2	0.461	0.456
Residual Std. Error	2.979 (df = 73)	2.992 (df = 71)
F Statistic	17.433*** (df = 4; 73)	11.756*** (df = 6; 71)

Note: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Table 12: Regression Results for mean_g_i on FS subset: Poor players

	(1)	(2)
alpha	0.059 (0.051)	0.057 (0.051)
beta	0.777 (0.601)	0.785 (0.601)
ineq_levellow	1.245** (0.414)	1.301** (0.418)
trtbaseline	-2.624*** (0.431)	-2.653*** (0.435)
genderMasculin		-0.283 (0.417)
NLE_score		0.866 (0.796)
Constant	4.200*** (0.562)	2.004 (2.230)
Observations	155	155
R2	0.283	0.291
Adjusted R2	0.264	0.262
Residual Std. Error	2.573 (df = 150)	2.576 (df = 148)
F Statistic	14.822*** (df = 4; 150)	10.127*** (df = 6; 148)

Note: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

CEE-M Working Papers[1] - 2024

WP 2025-01 **Katrin Erdlenbruch**, **Mabel Tidball** & Julia de Frutos Cachorro« Resource extraction and land-use choice in a two-player two-period game »

WP 2024-02 Gabriel Bayle & **Marc Willinger**
« Efficiency of the Minimum Approval Mechanism with heterogeneous players»

[1] **CEE-M Working Papers / Contact** : laurent.garnier@inrae.fr

- RePEc https://ideas.repec.org/s/hal/wpceem.html
- HAL https://halshs.archives-ouvertes.fr/CEE-M-WP/