Sustainability as a dynamic game*

Berno Buechel[†, ♯] Corinne Dubois[♯], Stephanie Fuerer[♯], Tjaša Maillard-Bjedov[♯]

[♯]Department of Economics, University of Fribourg, 1700 Fribourg, Switzerland

September 15, 2022

Abstract

Sustainability is a fundamental concept in the environmental domain, but also in other domains such as financial matters or regarding personal health. Sustainability means using resources today in a way that does not compromise the availability of resources tomorrow. In this paper, we propose and test a model that incorporates the essential features of sustainability. First, our Sustainability Game is dynamic in the sense that the actions played in each period have consequences for future periods. Second, there is a contribution threshold that must be reached in order to keep the level of resources. Finally, as in many applications, including environment conservation, cooperation between many individuals is required to reach a sustainable path, while the temptation to over-use resources is strong. We first derive equilibrium behavior analytically and then test these predictions in a lab experiment. Our main results are the following: (i) Theoretically and experimentally, strategic interaction reduces cooperative behavior and undermines sustainability. (ii) Theoretically and experimentally, lowering the threshold fosters cooperation and sustainability. Moreover, cooperative behavior correlates positively with a participant's agreeableness, pro-environmental attitudes and, in situations of equilibrium selection, with cognitive ability. Our results suggest that technological advancements that lower the threshold for sustainability and behavioral change toward sustainability need not be viewed as alternatives, but rather as complementary.

Keywords: Sustainability, dynamic game, common pool resource, public goods game, social dilemma, threshold, cooperation

JEL Classification Codes: C73, C92, H41, Q56

*We gratefully acknowledge funding from the University of Fribourg Research Pool (decision of 2020-03-20). We thank Dominik Karos and Holger Herz for their very helpful comments.

[†]Corresponding author. University of Fribourg, Department of Economics, Bd. de Pérolles 90, CH-1700 Fribourg, Switzerland. Email address: berno.buechel@unifr.ch, web: www.berno.info. Declarations of interest (all authors): none.

1 Introduction

Sustainability has become the cornerstone of many policy goals, project plans and personal decisions. The United Nations define sustainability as "meeting the needs of the present without compromising the ability of future generations to meet their own needs". In other words, sustainability means using resources today in a way that does not compromise the availability of resources in the future. We develop a model to study sustainability theoretically and experimentally. Our Sustainability Game has three essential features that make it suitable for studying this concept. First, it is dynamic so that decisions made today have an impact on the availability of resources in the future. Second, it includes threshold effects, such that a limited amount of resource utilization can be absorbed and has no impact on the future, but an over-use (i.e., usage above the threshold) leads to an irreversible decline in future resources. Third, the game features a tension between private and collective interests when there are multiple decision makers. While the literature provides examples of dynamic and threshold public good games, our contribution is to bring these elements together in a tractable model for analyzing sustainability.

We first derive theoretical equilibria and predictions from this Sustainability Game. We then test these predictions in a laboratory experiment. We focus on varying two dimensions: the number of decision makers and the sustainability threshold. In the experiment, we divide participants into three treatment groups. In the baseline treatment (*T-Baseline*) there are four decision makers, while in the *T-OnePlayer* treatment there is only one. Comparing the two treatments allows an assessment of the extent to which free-riding incentives impact sustainability. In many applications, such as climate and environment preservation, cooperation between many individuals is required to reach a sustainable path, and the temptation to free ride and over-use resources is strong. There are however some instances where the individual is solely responsible of maintaining a resource. One such example is personal health. An individual can refrain from unhealthy habits such as smoking, drinking, or eating sugary or fatty products and thereby maintain a better health level. It is therefore interesting to analyze how an individual manages sustainability and compare it to the case of multiple decision makers. Our results (both theoretical and experimental) indicate that players are more likely to reach the sustainability threshold when they are solely responsible for the decision; they are also less likely to choose defection (i.e., to contribute zero), they contribute a larger share of their endowment to the public account and manage to sustain a higher endowment throughout the game.

We then analyze how varying the sustainability threshold affects the strategies chosen by players. The treatment *T-LowThreshold* features a lower sustainability threshold than *T-Baseline*, which should theoretically lead to more cooperation (i.e., contributing according to the threshold), as mutual cooperation becomes an equilibrium. Our experimental results confirm the theoretical hypothesis: a lower threshold increases cooperation and sustainability and reduces defection. In the context of climate change, lowering the threshold could be interpreted as an improvement of technology to absorb CO2 emissions, for example. Such a technology would lower the effort required by the population to actually meet a sustainability threshold and could in principle make cooperation easier to sustain. Technological advancements and people's efforts to reduce consumption could therefore be complements rather than substitutes.

Finally, we investigate how specific personal characteristics affect the choice of strategies in the Sustainability Game. We find that agents who score higher on agreeableness tend to contribute more to the public account, in particular when contributing zero is the only equilibrium strategy. Our results also show that agents with higher cognitive abilities more often play equilibrium strategies. Similarly to Proto et al. (2019), we find that in situations when there are several feasible equilibria, participants with high cognitive ability are more likely to choose the socially optimal equilibrium. Finally, when controlling for agreeableness and cognitive ability and further personality traits, cooperation in our Sustainability Game still correlates positively with pro-environmental orientation yielding suggestive evidence for the external validity of our setup.

In experimental economics, social dilemmas are commonly studied by means of a public goods (PG) game. This strand of the literature mostly focuses on a *static* environment to test cooperation and self-interested behavior under different assumptions. The general result is that individuals fail to fully cooperate and under-contribute to the collective good even when interactions are repeated. For a review of the literature, see Ledyard (1995), Chaudhuri (2011) and Dal Bó and Fréchette (2018). A static environment, be it repeated or not, is however insufficient for analyzing sustainability. Dynamics is essential: the chosen actions today, if unsustainable, affect the possible actions in the future. Moreover, most PG games do not embody a threshold that distinguishes sustainable behavior from unsustainable behavior, the other essential aspect of sustainability.[1]

Our paper is also related to the literature on the commons, including Ostrom (1990), Levhari and Mirman (1980) and Fudenberg and Tirole (1991). Our setup differs from the classic common pool resource (CPR) game in several ways. First and foremost, our game includes threshold effects while the CPR game features a continuous growth, respectively decay, of the pool. This means in particular that a CPR game does not include the possibility that different small levels of usage are absorbed in the sense that they do not affect the size of the resource. The CPR is mostly used for studying renewable resources such as fisheries or forests. Our setup better represents situations in which threshold effects are important. In the case of global warming, for example, a certain level of CO2 emissions can be absorbed by the earth. Experts argue that it is only above the threshold that negative consequences occur. Second, in our game, the depletion of resources due to over-use is irreversible. This feature is well-suited to study environmental issues. In the context of global warming experts predict some adverse consequences of global warming that are irreversible, such as permafrost thaw, the increase in the levels of the oceans or species extinction (see Portner et al., 2022).[2]

More recently, dynamic PG and CPR games are being considered in the literature. Battaglini et al. (2016) study free-riding incentives in a durable PG game and compare the evolution of the durable public good when investment in it is reversible or irreversible. Vespa (2020) analyze the selection of strategies in a dynamic CPR game. They show that the Markov Perfect Equilibrium (MPE) is the modal strategy in this context. Our study also employs MPE as an equilibrium concept.[3] Our theoretical framework essentially differs from Battaglini et al. (2016) and Vespa (2020) in its introduction of a threshold. Moreover, the experimental design in Vespa (2020), and in many other studies on PG games, restricts participants' choice to a predetermined set of strategies while we allow them to choose any feasible contribution level. This approach may induce more noise in the data, but ensures that the emergence of an equilibrium strategy is truly the choice of participant and not partially imposed by the design.

The remainder of this paper is organized as follows. In Section 2, we present the theoretical model and characterize its equilibria. In Section 3, we describe the experimental design. In Section 4, we report the experimental results. Section 5 concludes.

[1]The sub-literature on threshold PG games (e.g. Cadsby and Maynes, 1999) builds a notable exception. These models are better suited to address sustainability, but are not truly dynamic either.

[2]Irreversible damage can also occur when acting unsustainably in the domains of health or financial matters.

[3]MPE is an advantageous equilibrium concept as it reduces the number of equilibria in dynamic settings. Its limitation is that it does not admit history-dependent strategies.

2 The Model

Our Sustainability Game features n agents who interact for an infinite number of periods. At the beginning of every period $t = 0, 1, 2, ...,$ each player $i = 1, 2, ..., n$ receives an endowment e_t. Each player i must then decide what share $c_{i,t}$ thereof she contributes to the public account; the remaining part of the endowment, $(1 - c_{i,t})e_t$, goes to her private account. After each period t, the game continues to period $t + 1$ with probability $\delta \in (0, 1)$, and ends with probability $(1 - \delta)$.

If the game continues to period $t + 1$, the total amount put on the public account by all n players in period t determines the endowment in period $t + 1$ in the following way:

$$e_{t+1} = \begin{cases} e_t & \text{if } \sum_{i=1}^{n} c_{i,t}e_t \geq Z_t \\ e_t - g\left(Z_t - \sum_{i=1}^{n} c_{i,t}e_t\right) & \text{else,} \end{cases} \tag{1}$$

where $Z_t = zne_t$ is the sustainability threshold set by a parameter $z \in [0, 1]$, and $g \in (0, \frac{1}{zn}]$ is a loss parameter. In words, the sustainability threshold parameter z defines which fraction of the overall endowment needs to be contributed to the public account. If the group as a whole contributes the sustainability threshold or more, the endowment in the next period will be identical to current endowment. If the sustainability threshold is not met, the next endowment will decline proportionately to the shortfall, according to loss parameter g. Note that the initial endowment e_0 is exogenous but endowments in all subsequent periods depend on players' decisions in past periods.

If the game ends, the payoffs are realized. Each player i's payoff consists of the sum of the amounts she kept in her private account in all periods up to the end of the game in period T, $P_i^T = \sum_{t=0}^{T}(1 - c_{i,t})e_t$.

Two specific characteristics of the game are worth mentioning. First, since the sustainability threshold is proportional to endowment, the situation is the same in relative terms for any endowment level and at any time. Second, the reduction in endowment, if it happens, is irreversible. A special case occurs when $g \equiv \frac{1}{zn}$, as contributing zero by all players then fully depletes the next endowment, i.e., $e_{t+1} = 0$.

2.1 The social planner's solution

Imagine a benevolent social planner could decide which fraction c_t of the aggregate endowment ne_t the whole group contributes to the public account in every period t. This social planner maximizes the aggregate expected payoff $EP_0 = \sum_{t=0}^{\infty} \delta^t \left(ne_t - c_t ne_t\right)$ over c_t. The solution to the social planner's problem delivers the socially optimal allocation of the Sustainability Game.

Proposition 1 (Social Planner's Solution)**.** *Suppose $\delta > \underline{\delta} := \frac{1}{1+ng(1-z)}$. The socially optimal contribution is $c_t^* = z$ in every period t.*

This proposition is proven in Appendix A.1. It means that if the probability of reaching the following period is sufficiently high, contributing the threshold – and thereby investing in future endowments without wasting contributions – maximizes social welfare. We are interested in problems where it is worthwhile from a social perspective to behave in a sustainable manner and therefore mostly restrict our analysis to cases where $\delta > \underline{\delta}$.

2.2 Markov perfect equilibria

To study individually rational contributions in the Sustainability Game described above, we look at symmetric Markov-perfect equilibria (MPE). MPE are a subset of sub-game perfect equilibria in which agents use stationary Markov strategies. Markov strategies do not depend on past decisions taken in a game, other than through the current levels of the state variables. In our game, the only state variable is the endowment. Each player thus maximizes her expected payoff conditional on the state variable e_t, which evolves according to Equation 1, and taking the actions of other players as given:

$$\max_{c_{i,t}} EP_{i,0} = \sum_{t=0}^{\infty} \delta^t \left(e_t - c_{i,t} e_t\right). \tag{2}$$

Proposition 2 (Symmetric Markov Perfect Equilibria)**.**

1. *If* $\delta < \bar{\delta} := \frac{1}{1+g(1-z)}$*, contributing zero (*Defection*) is the unique symmetric MPE.*
2. *If* $\bar{\delta} \leq \delta \leq \bar{\bar{\delta}} := \frac{1}{1+g(1-zn)}$ *there are three symmetric MPE (for* $n > 1$*):*
   i. *contributing* $c_{i,t} = 0$ *(*Defection*),*
   ii. *contributing* $c_{i,t} = z$ *(*Cooperation*),*
   iii. *contributing* $c_{i,t} = \frac{1-\delta(1-g(zn-1))}{\delta g(n-1)} \in [0, z]$ *(*Inbetween*).*
3. *If* $\delta > \bar{\bar{\delta}}$ *, contributing* $c_{i,t} = z$ *(*Cooperation*) is the unique symmetric MPE.*

The proposition is proven in Appendix A.2. It shows that for low discount factor δ the Sustainability Game is a social dilemma situation, in which individual incentives are to contribute zero (*Defection*), while the collective optimum is achieved with contribution of z (*Cooperation*) by Proposition 1. For a high discount factor, *Cooperation* is the unique equilibrium and for an intermediate discount factor both behaviors are symmetric MPE, while there is an additional equilibrium with intermediate contributions. When there is a multiplicity of equilibria, we consider *Defection* and *Cooperation* as focal and *Inbetween* as non-focal, a speculation that will be checked empirically. For the special case that there is only a single player ($n = 1$), we have $\underline{\delta} = \bar{\delta} = \bar{\bar{\delta}}$. Then the individually and socially optimal behavior is *Cooperation* (*Defection*) for a discount factor above (below) these thresholds.

Note that in the Sustainability Game, players' contributions are typically strategic complements as long as the aggregate contribution lies below the threshold. As soon as the threshold is reached, the contributions of different players become strategic substitutes.

3 Experimental Design

The experiment consists of two parts. In the first part, participants play the Sustainability Game six times. We use a between-subject design where we randomly assigned 18 sessions to three treatments called *T-Baseline*, *T-OnePlayer*, and *T-LowThreshold*. The treatments differ from one another with respect to (i) the number of players and (ii) the sustainability threshold, as explained below. The second part of the experiment is identical for all treatments and used to elicit personal characteristics of the players, including cognitive ability, agreeableness, risk aversion, ecological attitudes and socio-demographic characteristics.

3.1 Baseline treatment *T-Baseline*

Participants are matched into groups of four players ($n = 4$). Each player receives an initial allocation of 100 points in the first period. The sustainability threshold parameter z is set to 0.5 and the loss parameter g is also set to 0.5. According to this parametrization of g, z and n in

T-Baseline the endowment evolution of Equation 1 simplifies to $e_{t+1} = \min\{e_t, C_t/2\}$, where C_t is the aggregate contribution of the group in period t. After each period players receive information about the contributions of the other members of their group, the aggregate contribution of their group, whether the threshold was met and their endowment for the next period.

The continuation probability is $\delta = 0.65$. Technically, the game is played in blocks of five periods and when the end of the game is drawn within a block, the participants are informed after the block. This allows comparability between groups, since all of them play for at least five periods in every game.[4] At the end of the game, players are informed about their gains in the game. Every group plays the Sustainability Game six times. Finally, one game is randomly selected to be payoff relevant.

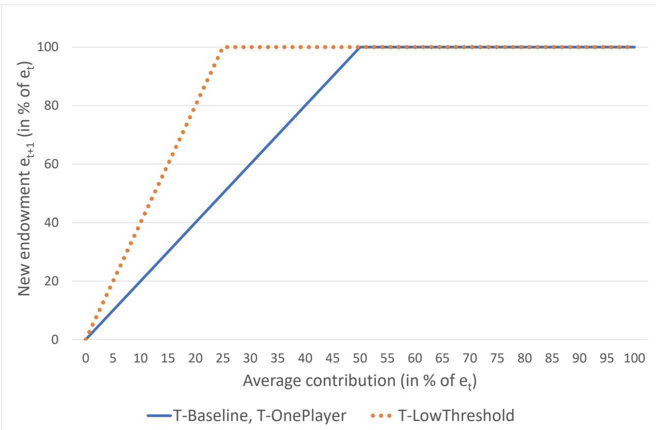
According to the parametrization in *T-Baseline*, we have $\underline{\delta} = 0.50$ and $\bar{\bar{\delta}} = 0.80$, which implies $\underline{\delta} < \delta < \bar{\bar{\delta}}$. By Propositions 1 and 2i., *T-Baseline* therefore represents a pure social dilemma: it is socially optimal for the group to contribute the threshold amount but it is individually rational for each player to defect, i.e., contribute zero.

3.2 No strategic interactions treatment *T-OnePlayer*

T-OnePlayer eliminates strategic interactions by setting the group size to $n = 1$, but it imitates *T-Baseline* in terms of the relationship between average contribution and future endowment (see Figure 1). For this purpose, the sustainability threshold parameter is kept at $z = 0.5$ and the loss parameter is set to $g = 2$. The evolution of endowments now reduces to $e_{t+1} = \min\{e_t, 2C_t\}$. In all other respects, the treatments are identical. In particular, there is an initial allocation of 100 points and the continuation probability is $\delta = 0.65$.

Under *T-OnePlayer* we have $\underline{\delta} = \bar{\delta} = \bar{\bar{\delta}} = 0.5 < \delta$. By Propositions 1 and 2i., *T-OnePlayer* therefore features a unique equilibrium that coincides with the social optimum: *Cooperation* (i.e., contributing according to the threshold).

Figure 1: Next endowment as a function of current average contribution



Notes: This figure shows how next period individual endowment e_{t+1} depends on the average contribution to the public account by the n group members. Both variables are expressed as percent of current individual endowment e_t. This graph applies to any period t, because in relative terms, the stage game is the same in every period.

[4]This approach, called block random design, follows Fréchette and Yuksel (2017).

3.3 Low-threshold treatment ***T-LowThreshold***

T-LowThreshold keeps group size $n = 4$ of the baseline, but lowers the threshold. Specifically, the sustainability threshold parameter is set to $z = 0.25$ and the loss parameter to $g = 1$. This parametrization reduces the threshold while keeping $g \equiv \frac{1}{zn}$, which ensures that a group that contributes zero to the public account has zero endowments in future periods. Figure 1 shows that in *T-LowThreshold* it is sufficient to reach an average contribution share of 25% of the endowment to maintain future endowments at the current level, whereas in *T-Baseline* (and in *T-OnePlayer*) an average contribution share of 50% is necessary. The expression for the evolution of endowments, Equation 1, reduces to $e_{t+1} = \min\{e_t, C_t\}$ in *T-LowThreshold*.

Under the parametrization of *T-LowThreshold* we have $\underline{\delta} = 0.25$, $\bar{\delta} \approx 0.57$ and $\bar{\bar{\delta}} = 1$ which implies $\underline{\delta} < \bar{\delta} < \delta = 0.65 < \bar{\bar{\delta}}$. By Propositions 1 and 2ii., *T-LowThreshold* therefore represents a problem of equilibrium selection, where both *Defection* and *Cooperation* are MPE, while *Cooperation* is still the social optimum. Table 1 summarizes the three treatments and their respective implications for the equilibria of the game.

Table 1: Treatments summary

	T-Baseline	*T-OnePlayer*	*T-LowThreshold*
group size n	4	1	4
sutainability threshold param. z	0.5	0.5	0.25
loss parameter g	0.5	2	1
product nzg	1	1	1
discount factor δ	0.65	0.65	0.65
social optimum	*Cooperation*	*Cooperation*	*Cooperation*
equilibria	*Defection*	*Cooperation*	*Cooperation*, *Defection*, *Inbetween*

Notes: *Cooperation* is defined as contributing according to the threshold, $c_{i,t} = z$. *Defection* is defined as contributing zero, $c_{i,t} = 0$. We restrict attention to symmetric Markov-perfect equilibria.

3.4 Personal characteristics

In the second part of the experiment, we elicit information about various characteristics of the participants. First, participants complete a risk aversion game. Players receive 30 points and must decide how many to invest in a profitable but risky project. We measure risk tolerance with the amount invested by each player. Then, we measure cognitive abilities of participants using 12 questions out of the Set 2 of the Raven Advanced Progressive Matrices. The number of questions a participants answers correctly gives her/his *Raven* score. Participants then take a personality test that consists of 24 items from the International Personality Item Pool, based on Maples-Keller et al. (2019). We measure two aspects of participants' personality: agreeableness and conscientiousness. Agreeableness measures a person's altruism, trust in others, cooperation and morality. Conscientiousness measures self-discipline, efficiency, achievement-striving and dutifulness. We then use the New Ecological Paradigm (NEP) of Dunlap et al. (2000) to assess pro-environmental orientation of participants.[5] After that, participants are asked to complete a short CO2 footprint questionnaire that consists of six questions from the WWF Swiss footprint calculator, following Berger and Wyss (2021). Our study finishes with a standard short demographic questionnaire.

[5]The New Ecological Paradigm Scale is a revised and extended version of the original New Environmental Paradigm Scale, also abbreviated with NEP.

3.5 Implementation

We programmed the experiment using the o-Tree framework of Chen et al. (2016), and ran the sessions online on the Prolific platform between January 19th and January 25th 2022. We ran a total of 18 sessions (6 per treatment), with a minimum of 4 and a maximum of 7 groups or single players per session. The assignment of sessions to treatments was randomly drawn by the computer. Participants were provided with detailed instructions for each part of the experiment and we ensured their understanding of the Sustainability Game with a comprehension questionnaire.[6] The study duration was around 60 minutes.[7] A total of 282 participants completed the study and earned on average 7.80 GBP.

3.6 Hypotheses

Before running the experiments, we had pre-registered four primary hypotheses and two secondary hypotheses.[8] The primary hypotheses concern treatment effects (derived from the game-theoretic equilibrium analysis); the secondary hypotheses concern expected correlations between individual traits and behavior.

Let us begin with the primary hypotheses about treatment *T-OnePlayer*. In the baseline setting, *T-Baseline*, *Defection* is the unique equilibrium. In contrast, in the setting without strategic interaction, *T-OnePlayer*, *Cooperation* is the unique equilibrium. Hence, we predict:

Hypothesis 1 (H-CoopDef-One)**.** *Without strategic interaction (i.e., in* T-OnePlayer*),* Cooperation *is more often played and* Defection *is less often played (than in the baseline,* T-Baseline*).*

This game-theoretic prediction captures free-riding incentives that arise when there are multiple players. Since *Cooperation* means reaching the threshold, while *Defection* does not, the behavior predicted in Hypothesis H-CoopDef-One has the following direct consequence:

Hypothesis 2 (H-Reached-One)**.** *Without strategic interaction (i.e., in* T-OnePlayer*), the threshold is reached more often (than in the baseline,* T-Baseline*).*

We now turn to treatment *T-LowThreshold*, which differs from the baseline by its lower threshold. In *T-LowThreshold* both *Defection* and *Cooperation* are equilibria. Thus, participants in that treatment might more often play *Cooperation* than in the baseline *T-Baseline*, where *Defection* is the unique equilibrium. Hence, we hypothesize:

Hypothesis 3 (H-CoopDef-Low)**.** *When the threshold is lower (i.e., in* T-LowThreshold*),* Cooperation *is more often played and* Defection *is less often played (than in the baseline,* T-Baseline*).*

This hypothesized behavior on the individual level has the following consequences on the collective level:

Hypothesis 4 (H-Reached-Low)**.** *When the threshold is lower (i.e., in* T-LowThreshold*), it is reached more often (than in the baseline,* T-Baseline*).*

In addition to these four primary hypotheses we also pre-registered two secondary hypotheses. First, the standard index of *Agreeableness* that we use includes measures of altruism, cooperation, trust and morality. These aspects of personality should translate into more pro-social behavior in our game and more specifically correlate positively with contributions. Hence, we hypothesize:

[6]The instructions, the questionnaire and all materials of the study design including the pre-registered hypotheses can be found at `https://doi.org/10.1257/rct.6132`.

[7]Some potential participants did not complete the study. Mostly, they dropped out at a very early stage.

[8]These hypotheses were pre-registered at AEAregistry-6132. They are copied and pasted here with identical wording and order. We here only add the treatment names in brackets, replacing longer explanations between the hypotheses in the 'Analysis plan' document.

Hypothesis 5 (H-Agree). *Agreeable people contribute more.*

Second, people with high cognitive ability might play more often the game-theoretic equilibrium strategies than people with lower cognitive ability. The intuition behind this hypothesis is that high cognitive ability individuals could act more strategically: they might use a deeper level of reasoning when taking into account their own and the other players' incentives. Hence, we expect participants with high *Raven* score to play more often *Cooperation* (in *T-OnePlayer* and *T-LowThreshold*) and to play more often *Defection* (in *T-Baseline* and *T-LowThreshold*). This is summarized by Hypothesis H-Raven.

Hypothesis 6 (H-Raven). *People with high cognitive ability play more often* Cooperation *and play more often* Defection.

4 Experimental Results

Before testing the six hypotheses, we briefly describe the data set.

4.1 Descriptive statistics

We have 282 participants in our study. They are aged between 19 and 63, with an average of around 27 years. 52% of participants declare themselves as women and 48% as men or other. On average our participants have a *Raven* score of 6.5 out of 12 with substantial variation between participants, and an agreeableness index of 45.0 (on a scale from 12 to 60). These and further descriptive statistics are summarized in the first block of Table 2.

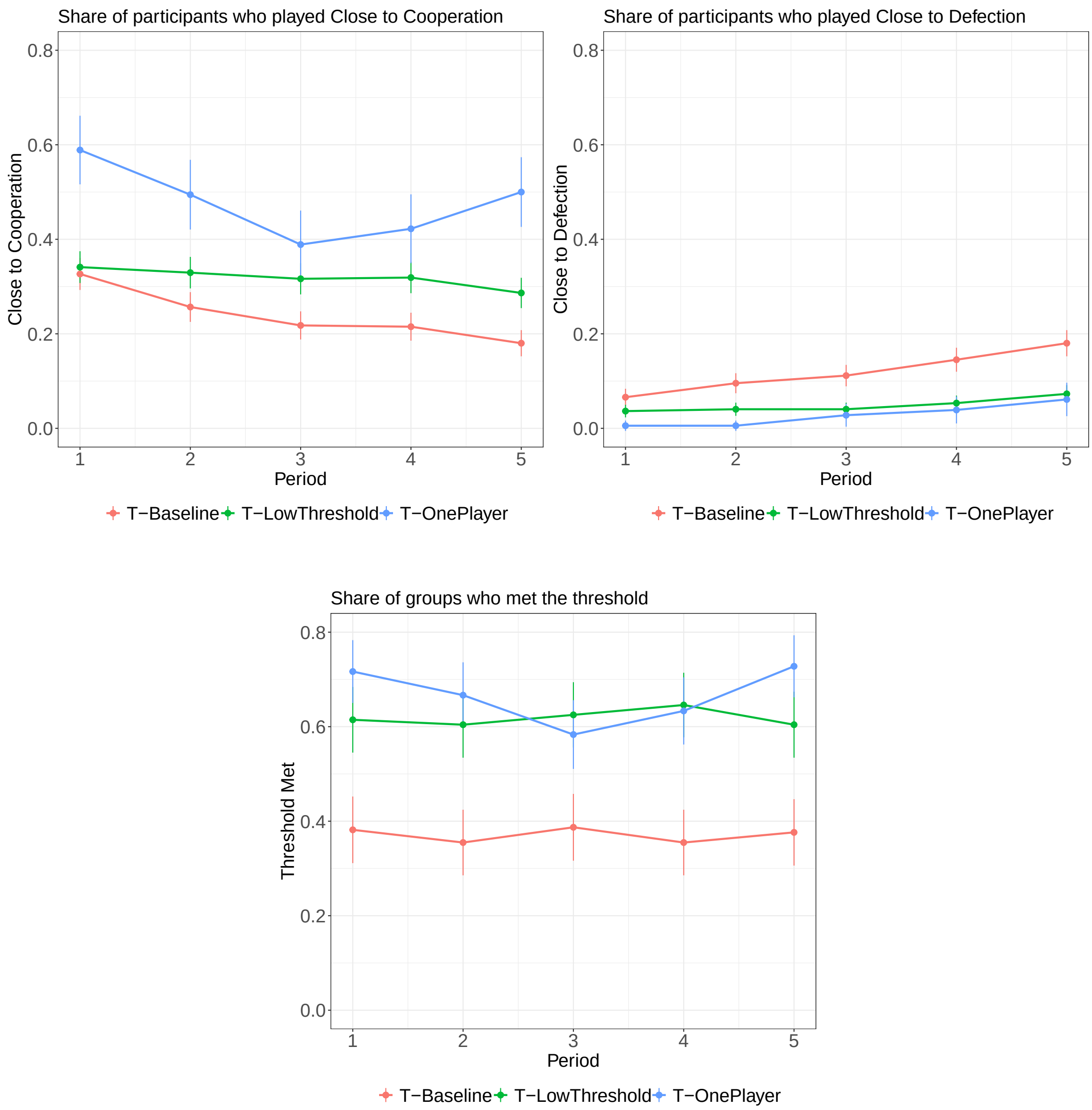
The participants form 63 groups of four (of which 31 are in treatment *T-Baseline* and 32 in treatment *T-LowThreshold*) and 30 remain single players (those in *T-OnePlayer*). Table 2 summarizes all important outcome variables both on the individual level and on the collective level. Each game lasts at least five periods and is repeated six times, yielding 2,790 observations on the group level and 8,460 observations on individual behavior.[9] The binary variables *Close to Cooperation*, *Close to Defection*, and *Close to Inbetween*, equal 1 if a participant has submitted a contribution share within a $\pm 2pp$ band around the strategy.[10] Pooling all individual decisions, participants play *Close to Cooperation* in 30.1% of all cases, while they play *Close to Defection* in 7.8% of all cases. The strategy *Close to Inbetween* can only be played in *T-LowThreshold* and is chosen in only 3.2% of these cases. *Threshold met* is a binary variable that equals 1 if the participant's group has reached the sustainability threshold in a given period t and zero otherwise. On average, participants reach the sustainability threshold 55% of the time. *Sustainability* is a binary variable that equals 1 if the group reached the threshold in all periods until t, or equivalently, maintained their endowment at its original level of 100.

Figure 2 illustrates the evolution of the main variables over time separated by treatment. In particular, it displays the percentage of participants who played *Close to Cooperation* (upper left panel) and *Close to Defection* (upper right panel), as well as the *Threshold met* (lower panel). Figure 2 yields a first impression of potential treatment effects.

[9] We focus on the first five periods in each game because these can be observed independent of the actual end of the game, thanks to the block random design.

[10] For *T-Baseline* and *T-OnePlayer Close to Cooperation* therefore equals 1 for contribution shares between 0.48 and 0.52 and 0 otherwise. For *T-LowThreshold Close to Cooperation* equals 1 for contribution share between 0.23 and 0.27 and 0 else. *Close to Defection* equals 1 if a participant chose a contribution share between 0 and 0.02.

Figure 2: Main variables evolution over time, by treatment



Notes: Mean and standard 95% confidence intervals, pooling groups and repetitions of the game. *Close to Cooperation* and *Close to Defection* are binary variables that equal 1 if a participant has submitted a contribution share within a $\pm 2pp$ band around the *Cooperation* and *Defection* strategies, respectively. *Threshold met* is a binary variable that equals 1 if a participant's group has reached the sustainability threshold.

Table 2: Summary statistics

Variable	N	Mean	St. Dev.	Min	Max
Female	282	0.521	0.500	0	1
Age	282	27.372	7.784	19	63
Raven	282	6.518	3.113	0	12
Risk Tolerance	282	14.372	7.403	0	30
Agreeableness	282	44.993	6.380	26	58
Conscientiousness	282	43.323	7.012	24	60
Contribution	8,460	29.549	19.747	0	100
Contribution Share	8,460	0.354	0.217	0.000	1.000
Private account	8,460	51.982	23.724	0	100
Close to Cooperation	8,460	0.301	0.459	0	1
Close to Defection	8,460	0.078	0.268	0	1
Close to Inbetween	3,840	0.032	0.177	0	1
Endowment	2,790	81.904	27.362	0	100
Threshold met	2,790	0.551	0.497	0	1
Sustainability	2,790	0.349	0.477	0	1

Notes: The variable *Raven* measures cognitive ability of the participants and corresponds to the number of questions that they answered correctly in the *Raven* test that consisted of 12 questions. *Risk Tolerance* is the number of points out of 30 invested in the profitable risky project in the risk aversion game, where a higher score indicates higher tolerance to risk. *Agreeableness* and *Conscientiousness* report participants' scores on the International Personality Item Pool test. Observations on individual behavior and on group outcomes are here pooled over all three treatments, all five periods, and all six repetitions of the game.

4.2 Test of primary hypotheses: treatment effects

This first part of our econometric analysis estimates treatment effects on the three main outcome variables, *Close to Cooperation* (Table 3), *Close to Defection* (Table 4), and *Threshold met* (Table 5). We use dummy variables for the treatments *T-LowThreshold* and *T-OnePlayer* and keep *T-Baseline* as the reference category. We run each regression once without and once with controls. The set of control variables is *Age*, *Female*, *Raven*, *Risk Tolerance*, *Agreeableness* and *Conscientiousness*. All regression models report robust standard errors clustered at the group level. The clustering method corrects for intra-group correlation of residuals (see Abadie et al., 2017).

Table 3 reports treatment effects on a participant's probability of playing *Close to Cooperation*. Columns (1) and (2) only consider the first period of each game, Columns (3) and (4) only consider period five. Recall that for all participants we have observations of their first five periods. The dependent variable is binary, so we estimate the coefficients using a logisitc regression model. In Columns (5) and (6), we take the individual's average of the binary variable *Close to Cooperation* over periods one to five of each game. The dependent variable therefore represents the frequency of playing *Close to Cooperation* over the five periods of the game; it is no longer binary but can take values $\in [0, 0.2, 0.4, 0.6, 0.8, 1]$. To easily interpret coefficients, we treat this dependent variable as continuous and run an OLS regression.

The results in Table 3 show that removing strategic interactions significantly increases the probability of playing *Close to Cooperation*. This effect is illustrated in the upper left panel of Figure 3, which shows the regression coefficients when running it for each of the five periods. We observe that the effect of removing strategic interaction occurs already in period one, is stable and persists until period five. Columns (5) and (6) of Table 3 indicate that this effect increases the frequency of playing *Close to Cooperation* by about 24-25pp. The effect is large and statistically significant. Note that the inclusion of the controls does not change the size or significance of the estimates. These results fully support Hypothesis H-CoopDef-One with respect to *Cooperation*.

The results in Table 3 also indicate that, in line with Hypothesis H-CoopDef-Low, lowering the threshold significantly increases the probability of *Close to Cooperation*, in period five, but not in period one. The left panel of Figure 3 confirms that treatment effects of *T-LowThreshold* increase over time and only become significant toward the end of the first five periods. Columns (5) and (6) of Table 3 suggest that lowering the threshold results in an increase of about 8pp in the frequency of playing *Close to Cooperation.* The weak significance of the *T-LowThreshold* coefficients in Columns (5) and (6) is due to the insignificance of treatment effects in early periods of the game. Hence, Hypotheses H-CoopDef-One and H-CoopDef-Low are largely supported concerning *Cooperation.*

Table 3: Treatment effects on playing Close to Cooperation

Dependent Variable:	Close to Cooperation					
	period 1		period 5		average 1-5	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
	Logit	Logit	Logit	Logit	OLS	OLS
Variables						
T-OnePlayer	1.083***	1.275***	1.516***	1.657***	0.2396***	0.2508***
	(0.3396)	(0.3407)	(0.3409)	(0.3518)	(0.0672)	(0.0617)
T-LowThreshold	0.0654	0.0814	0.6030**	0.6451**	0.0792	0.0821*
	(0.2575)	(0.2566)	(0.2849)	(0.2747)	(0.0510)	(0.0461)
Constant	-0.7235***	-1.676	-1.516***	-1.432*	0.2392***	0.1412
	(0.1889)	(1.045)	(0.2285)	(0.8541)	(0.0360)	(0.1545)
Controls	No	Yes	No	Yes	No	Yes
Fit statistics						
Observations	1,692	1,692	1,692	1,692	1,692	1,692
R^2					0.04257	0.13610
Pseudo R^2	0.01973	0.09267	0.03928	0.07353		
Wald (joint nullity)	5.7419	5.3471	9.9380	4.8813	6.3765	7.1612

Clustered (Group) standard errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

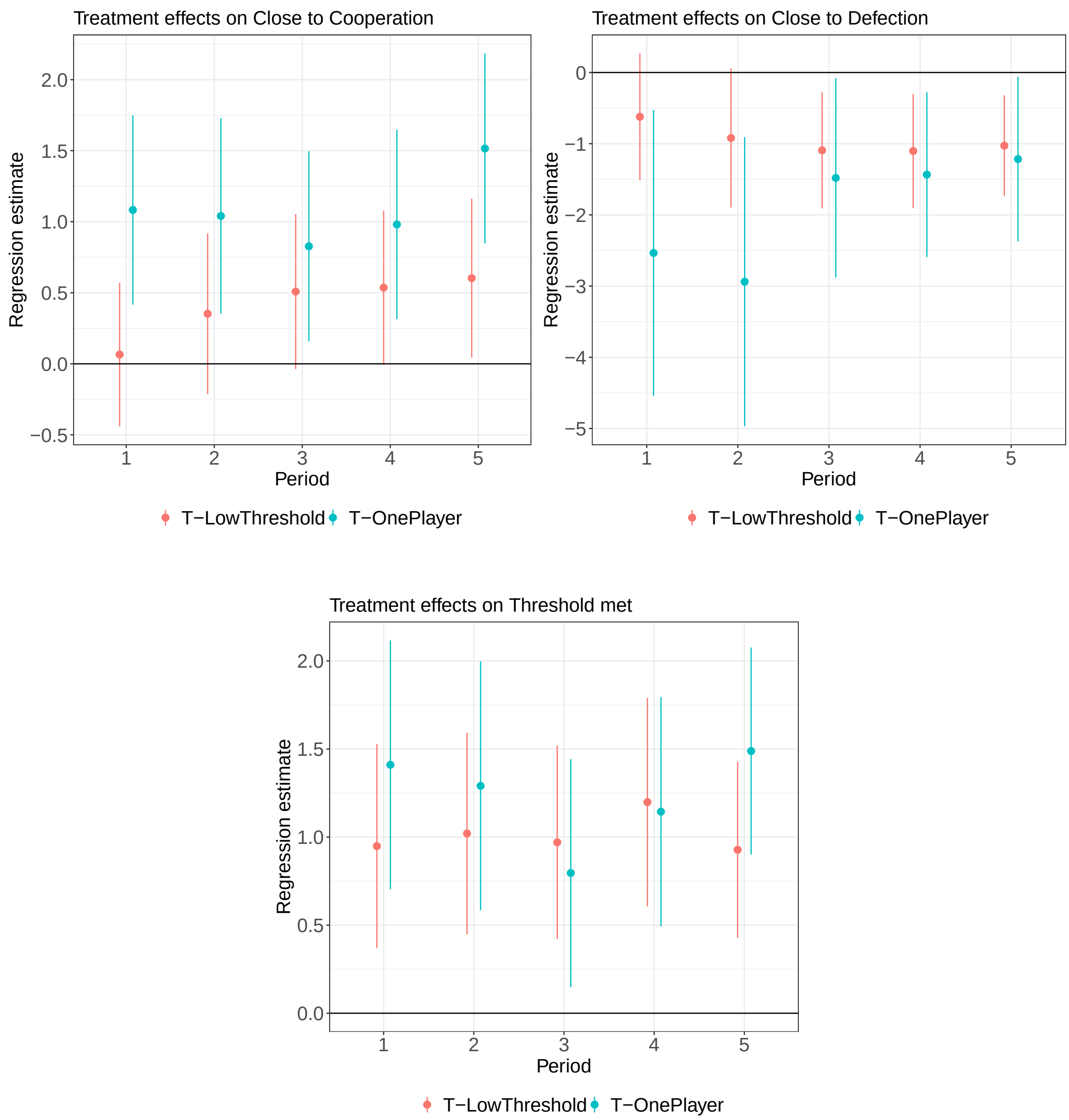
Concerning *Defection*, the right panel of Figure 3 illustrates the treatment effects on a participant's probability of playing *Close to Defection.* Table 4 shows the corresponding regressions. Lowering the threshold significantly reduces the probability of *Close to Defection*, again with stronger effect in period five than in period one. The average frequency of playing *Close to Defection* falls by about 8 to 9pp in *T-OnePlayer* and by 7pp in *T-LowThreshold* relative to *T-Baseline.* The results in Table 4 and Figure 3 show that Hypotheses H-CoopDef-One and H-CoopDef-Low are not only supported concerning *Cooperation* but also concerning *Defection.*[11]

The remaining primary hypotheses are Hypotheses H-Reached-One and H-Reached-Low, which predict that without strategic interaction, or with a lower threshold, the threshold is reached more often. Table 5 presents the results.[12] It shows that both *T-OnePlayer* and *T-LowThreshold* significantly increase the groups' probability of reaching the threshold in period one and period five. For all periods in between the coefficients are also significant, as illustrated in the lower panel of Figure 3. The average frequency of reaching the threshold increases by 29-33pp for *T-OnePlayer* and by 25-27pp for *T-LowThreshold* (Columns (5) and (6) of Table 5). The effects are large and

[11]Notice that for the strategy *Inbetween* there are no treatment effects to consider, as it is only playable in one treatment, *T-LowThreshold.*

[12]Note that the regression in Table 5 concern group-level data and not individual data, which is why the number of observations is smaller.

Figure 3: Treatment effects on strategies played, over time



Notes: Coefficient and 95% confidence intervals from logistic regression with dependent variable *Close to Cooperation* (upper left panel) and *Close to Defection* (upper right panel) and *Threshold met* (lower panel), on a constant, *T-LowThreshold* and *T-OnePlayer* dummies. Estimation as in Tables 3-5 with standard errors clustered on the group level.

Table 4: Treatment effects on playing Close to Defection

Dependent Variable:	Close to Defection					
	period 1		period 5		average 1-5	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
	Logit	Logit	Logit	Logit	OLS	OLS
Variables						
T-OnePlayer	-2.535**	-2.585**	-1.216**	-1.122*	-0.0919***	-0.0834***
	(1.024)	(1.083)	(0.5897)	(0.5889)	(0.0288)	(0.0294)
T-LowThreshold	-0.6224	-0.7173*	-1.027***	-1.069***	-0.0709**	-0.0703**
	(0.4533)	(0.3921)	(0.3612)	(0.3650)	(0.0284)	(0.0270)
Constant	-2.652***	3.297	-1.516***	2.137	0.1196***	0.3959***
	(0.2482)	(2.205)	(0.2322)	(1.382)	(0.0245)	(0.1098)
Controls	No	Yes	No	Yes	No	Yes
Fit statistics						
Observations	1,692	1,692	1,692	1,692	1,692	1,692
R^2					0.03533	0.07386
Pseudo R^2	0.02890	0.08459	0.03888	0.08263		
Wald (joint nullity)	3.6184	3.0670	5.0063	6.0010	5.1168	3.6988

Clustered (Group) standard errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

highly significant and hence fully support Hypotheses H-Reached-One and H-Reached-Low.

4.3 Implications of treatment effects

The treatment effects that we find should have implications for further outcome variables of interest, in particular for the evolution of endowments and for maintaining sustainability. Figure 4 illustrates the evolution of endowments and sustainability over time for each treatment. While in the baseline treatment only 10.8% of the groups act sustainable in the first five periods, this fraction increases to 33.9% when there is no strategic interaction and to 21.9% when the threshold is lower.

Table 6 estimates the treatment effects of removing strategic interactions and lowering the threshold on groups' *Sustainability* at the end of period five and on their endowments at period six.[13] Columns (1) and (2) indicate that *T-OnePlayer* and *T-LowThreshold* both have a positive impact on a group's probability of maintaining a sustainable behavior until the end of period five. Columns (3) and (4) reveal an increase in the period 6 endowment by 18-21 points for *T-OnePlayer* and by 27-28 points for *T-LowThreshold*, in comparison to *T-Baseline*. The effects are hence substantial. Finally, we estimate treatment effects on the time trend of *Endowment*. Appendix Table B.1 shows that in *T-Baseline*, *Endowment* falls by 11.4 points on average in every period. The decline in *Endowment* is significantly smaller in *T-OnePlayer* (7.5 points) and *T-LowThreshold* (5.3 points). To summarize, we find that that without strategic interaction and with a lower threshold, indeed, endowments remain higher and sustainability can be more often maintained.

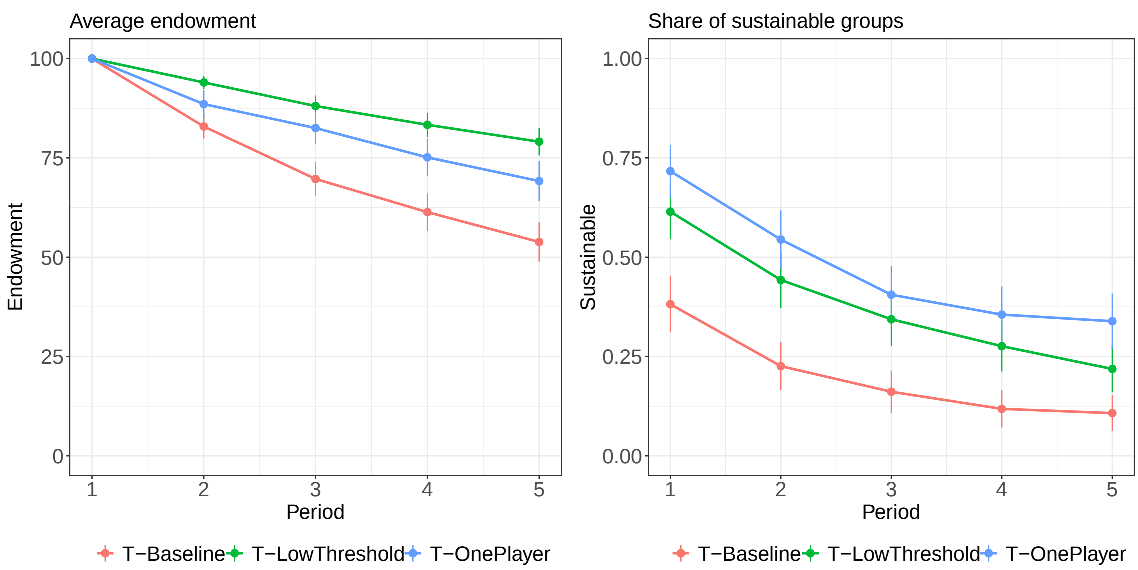
[13]The endowment at period six only depends on decisions in the first five periods. Recall that sustainability is defined as 1 if the endowment has never decreased, and 0 otherwise.

Table 5: Treatment effects on Threshold Met

Dependent Variable:	Threshold met					
	period 1		period 5		average 1-5	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
	Logit	Logit	Logit	Logit	OLS	OLS
Variables						
T-OnePlayer	1.410***	1.686***	1.488***	1.768***	0.2946***	0.3325***
	(0.3595)	(0.3665)	(0.2993)	(0.2995)	(0.0647)	(0.0597)
T-LowThreshold	0.9489***	1.124***	0.9280***	1.088***	0.2478***	0.2711***
	(0.2951)	(0.3217)	(0.2546)	(0.2796)	(0.0521)	(0.0536)
Constant	-0.4823**	2.958*	-0.5051***	4.477**	0.3710***	1.118***
	(0.2133)	(1.707)	(0.1630)	(1.740)	(0.0408)	(0.2725)
Controls	No	Yes	No	Yes	No	Yes
Fit statistics						
Observations	558	558	558	558	558	558
R^2					0.14110	0.19742
Pseudo R^2	0.05860	0.09264	0.06326	0.09412		
Wald (joint nullity)	9.1344	4.0404	14.443	5.5233	14.535	7.5775

Clustered (Group) standard errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Figure 4: Evolution of Endowments and Sustainability over time, by treatment



Notes: Mean and standard 95% confidence intervals, pooling groups and repetitions of the game. *Sustainability* is a binary variable that equals 1 if a participant's group has reached the sustainability threshold in all periods so far.

Table 6: Treatment effects on Sustainability and Endowment

Dependent Variables:	Sustainability period 5		Endowment period 6	
Model:	(1) Logit	(2) Logit	(3) OLS	(4) OLS
Variables				
T-OnePlayer	1.448***	1.593***	17.66**	21.11***
	(0.5007)	(0.5117)	(6.880)	(6.792)
T-LowThreshold	0.8433*	1.004**	26.71***	28.14***
	(0.4597)	(0.4731)	(5.406)	(5.653)
Constant	-2.116***	1.353	47.34***	91.80***
	(0.4112)	(2.048)	(4.504)	(28.99)
Controls	No	Yes	No	Yes
Fit statistics				
Observations	558	558	558	558
R^2			0.10391	0.15144
Pseudo R^2	0.05001	0.07411		
Wald (joint nullity)	4.2717	1.8266	12.214	4.3837

Clustered (Group) standard errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

4.4 Test of secondary hypotheses: agreeableness and cognitive ability

Hypothesis H-Agree stipulates a positive relation between *Agreeableness* and *Contribution Share*. In Table 7, we test this hypothesis by regressing *Contribution Share* on *Agreeableness*, treatment dummies and the set of control variables (*Age*, *Female*, *Raven*, *Risk Tolerance*, *Conscientiousness*). Columns (2), (4) and (6) additionally include the interaction of *Agreeableness* with treatment dummies to allow for heterogeneous effects of *Agreeableness* across treatments.[14]

Table 7 indicates that *Agreeableness* does not significantly affect *Contribution Share* in the first period of the game, while there is a significant effect in period five and on the average. A one-point increment in *Agreeableness* raises contribution shares by 0.36pp in period five (Column (3)) and by 0.26pp on average (Column (5)), a small effect.[15] The result suggests that participants who score high on *Agreeableness* maintain slightly higher contribution levels throughout the game.

Columns (4) and (6) indicate that *Agreeableness* has heterogeneous effects across treatments. The effect of *Agreeableness* is strongest in the baseline treatment. The effect is weaker in *T-LowThreshold* indicating that *Agreeableness* matters more in situations where *Cooperation* is more difficult, i.e., when it is not an equilibrium of the game. We also observe that the effect of *Agreeableness* disappears in *T-OnePlayer*, indicating that *Agreeableness* only matters when multiple players are involved in the game. Intuitively, participants with higher agreeableness score show more agreeable behavior toward others. Hypothesis H-Agree hence finds some support, but mainly in the pure social dilemma situation.

The final hypothesis, Hypothesis H-Raven, predicts that participants with high cognitive abilities, as measured by high *Raven* scores, more often play equilibrium strategies in the Sustainability Game. Tables 8 and 9 test the relation between *Raven* score and playing *Close to Cooperation* and *Close to Defection*, respectively.

[14] For treatment effects concerning the outcome variable *Contribution Share*, see Appendix B.2.

[15] The *Agreeableness* score ranges from 12 to 60 points with a standard deviation of 6.38. A participant whose *Agreeableness* is one standard deviation higher than another participant's would therefore contribute on average $6.38 \times 0.26 = 1.7pp$ more.

Table 7: Effect of Agreeableness on Contribution Shares (OLS)

Dependent Variable:	Contribution Share					
	period 1		period 5		average 1-5	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Agreeableness	0.0003	0.0003	0.0036***	0.0060***	0.0026***	0.0046***
	(0.0013)	(0.0025)	(0.0013)	(0.0019)	(0.0010)	(0.0015)
T-OnePlayer × Agree.		-0.0033		-0.0076**		-0.0069**
		(0.0039)		(0.0035)		(0.0030)
T-LowThreshold × Agree.		0.0009		-0.0040*		-0.0029
		(0.0027)		(0.0023)		(0.0019)
Constant	0.4500***	0.4469***	0.2891***	0.1566	0.3464***	0.2400***
	(0.0793)	(0.1334)	(0.0703)	(0.1029)	(0.0600)	(0.0840)
Treatment dummies	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	1,692	1,692	1,692	1,692	1,692	1,692
R^2	0.19261	0.19396	0.16136	0.16592	0.30834	0.31503
Wald (joint nullity)	28.954	25.023	24.533	26.462	38.345	38.567

Clustered (Group) standard errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Columns (1) and (3) of Table 8 report the effect of *Raven* on the probability of playing *Close to Cooperation* in period one and five, respectively. We control for treatments and the set of other control variables which is *Age*, *Female*, *Risk Tolerance*, *Agreeableness* and *Conscientiousness*. The estimate is positive and highly significant in both periods, thus confirming that more intelligent players more often play *Close to Cooperation*. Column (5) reports the effect of *Raven* on the average frequency of playing *Close to Cooperation* in the five periods of the game. It indicates that a one-point increase in *Raven* is associated with a 3pp increase in the frequency of playing *Close to Cooperation*. It follows that a participant whose *Raven* score is one standard deviation higher than another participant's would cooperate on average $3.11 \times 0.03 = 9pp$ more often.

However, it is important to remember that *Close to Cooperation* is not an equilibrium strategy in *T-Baseline*, but only in *T-OnePlayer* and *T-LowThreshold*. According to Hypothesis H-Raven, the effect of cognitive ability should therefore be stronger in *T-OnePlayer* and *T-LowThreshold* relative to *T-Baseline*. Columns (2), (4), and (6) examine the heterogeneity of the effect of *Raven* across treatments. In the baseline treatment *T-Baseline* the effect is weaker than overall. On the other hand, we observe a weakly significant increase in the effect of *Raven* on playing *Close to Cooperation* in *T-LowThreshold* relative to *T-Baseline* in period one and on average. Therefore, the effect of *Raven* on playing *Close to Cooperation* is mainly driven by *T-LowThreshold* where *Cooperation* is indeed an equilibrium. This observation is in line with the causal evidence provided by Proto et al. (2019) who show that participants with high *Raven* score are better able to select the Pareto-dominant equilibrium.

Concerning *Defection*, Columns (1), (3) and (5) of Table 9 indicate no significant influence of *Raven* on the probability of playing *Close to Defection*. According to Hypothesis H-Raven though, we only expect *Raven* to have an impact on playing *Close to Defection* when *Defection* is an equilibrium strategy, i.e., in treatments *T-Baseline* and *T-LowThreshold*. For *T-Baseline* Columns (2), (4) and (6) show a significant positive effect in period five and on average, but not

Table 8: Effect of Raven on playing Close to Cooperation

Dependent Variable:	Close to Cooperation					
	period 1		period 5		average 1-5	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
	Logit	Logit	Logit	Logit	OLS	OLS
Variables						
Raven	0.2058***	0.1165*	0.1317***	0.0717	0.0302***	0.0150
	(0.0434)	(0.0639)	(0.0344)	(0.0592)	(0.0063)	(0.0098)
T-OnePlayer × Raven		0.1420		0.0108		0.0241
		(0.1158)		(0.0993)		(0.0178)
T-LowThreshold × Raven		0.1587*		0.1126		0.0256*
		(0.0899)		(0.0795)		(0.0135)
Constant	-1.676	-0.8863	-1.432*	-0.8301	0.1412	0.2609*
	(1.045)	(1.044)	(0.8541)	(0.8730)	(0.1545)	(0.1471)
Treatment dummies	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	1,692	1,692	1,692	1,692	1,692	1,692
R^2					0.13610	0.14817
Pseudo R^2	0.09267	0.10015	0.07353	0.07753		
Wald (joint nullity)	5.3471	6.0530	4.8813	4.3861	7.1612	7.4095

Clustered (Group) standard errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

in period one. This suggests that high cognitive ability players do not necessarily play the unique equilibrium strategy more often from the beginning of the game, but settle for it more frequently later. Hence, if *Raven* positively affects the probability of *Close to Defection*, this is mainly driven by the treatment where *Defection* is the unique equilibrium strategy. Interestingly, in the treatment *T-LowThreshold*, in which both *Defection* and *Cooperation* are equilibria, this effect is nullified, again consistent with the idea that in a situation of equilibrium selection, high *Raven* participants might more often select the Pareto-efficient equilibrium.

4.5 Further results: waste and environmental attitudes

The test of the secondary hypotheses shows that there are associations between the personality traits agreeableness and cognitive ability with behavior in the Sustainability Game. In particular, in line with Hypothesis H-Raven high cognitive ability players more often play equilibrium strategies. As a corollary, we expect that they are less likely to choose "wasteful" strategies. *Waste* occurs when a group contributes to the public account more than the sustainability threshold. In Appendix B.3 we explore this relation between cognitive ability and *Waste*. Participants with higher Raven score indeed less often contribute above the threshold z (Appendix Table B.3) with the consequence that groups with higher average Raven score less often induce waste (Appendix Table B.4).

Finally, we explore how cooperative behavior in the Sustainability Game correlates with environmental attitudes. We find systematic correlations with the New Ecological Paradigm (NEP) score, but no relation with our small CO2 footprint score. The NEP measures the endorsement of a "pro-ecological" world view (Dunlap et al., 2000).[16] The results in Tables B.4 indicate that par-

[16]The NEP score ranges from 26 to 71 with mean 55.5 and standard deviation 7.7.

Table 9: Effect of Raven on playing Close to Defection

Dependent Variable:	Close to Defection					
	period 1		period 5		average 1-5	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
	Logit	Logit	Logit	Logit	OLS	OLS
Variables						
Raven	-0.0486	-0.0554	0.0727	0.1171**	0.0029	0.0101*
	(0.0615)	(0.0671)	(0.0488)	(0.0529)	(0.0035)	(0.0059)
T-OnePlayer × Raven		0.1547		0.0096		-0.0093
		(0.1007)		(0.1668)		(0.0074)
T-LowThreshold × Raven		0.0100		-0.1340		-0.0127*
		(0.1346)		(0.1051)		(0.0072)
Constant	3.297	3.368	2.137	1.644	0.3959***	0.3365***
	(2.205)	(2.071)	(1.382)	(1.359)	(0.1098)	(0.1046)
Treatment dummies	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	1,692	1,692	1,692	1,692	1,692	1,692
R^2					0.07386	0.08242
Pseudo R^2	0.08459	0.08511	0.08263	0.08750		
Wald (joint nullity)	3.0670	2.9632	6.0010	5.5413	3.6988	3.2550

Clustered (Group) standard errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

ticipants with a stronger ecological concern are more likely to play *Close to Cooperation* beyond the effects of other control variables (Table B.5). The effect is mainly driven by *T-OnePlayer*. Appendix Table B.6 shows that a higher NEP score is associated with a lower probability of playing *Close to Defection*, without any significant differences among the treatments. Therefore, sustainable behavior in our Sustainability Game is positively associated with pro-environmental orientation.

5 Discussion

Sustainability is defined as using resources today in a way that does not compromise the availability of resources in the future. This definition entails that there is a threshold that distinguishes sustainable behavior, which can be repeated, from unsustainable behavior, which cannot. In this paper we propose a model of sustainability. It differs from the standard public goods (PG) and common pool resource (CPR) games by being truly dynamic and having a threshold. Both these aspects are necessary to capture the meaning of sustainability, as defined above. In particular, our model incorporates a resource's capacity to absorb a certain level of consumption, while over-consumption (irreversibly) diminishes the resource.

Theoretically and experimentally, we find that cooperation is fostered by excluding interaction and by lowering the sustainability threshold. The former result reflects analogous fundamental findings for PG and CPR games, where making one player solely responsible eliminates the free-rider problem. The latter result goes beyond a mechanical effect of reaching a threshold more often because it is set lower. We find in particular that lowering the threshold makes people choose Defection less frequently. Drawing conclusions for the the goal of sustainability (be it in the environmental domain or concerning financial matters or personal health), our results suggests that technological progress need not be always seen as a substitute to behavioral change; if it lowers the threshold, it can also work as a complement that fosters behavioral change toward sustainability.

A Appendix: Proofs

A.1 Proof of Proposition 1: social planner's solution

Proposition A.1 below nests Proposition 1 of the main text as a special case.

Proposition A.1 (Social Planner's Solution, General)**.** *Consider a social planner that maximizes the aggregate expected payoff* $EP_0 := \sum_{t=0}^{\infty} \delta^t (ne_t - c_t ne_t)$ *over the fraction* c_t *of the total endowment that the group contributes.*

1. *If* $\delta > \underline{\delta} := \frac{1}{1+ng(1-z)}$*, the social planner chooses* $c_t^* = z$ *at each period* t*.*
2. *If* $\delta = \underline{\delta}$*, the social planner chooses* $c_t^* \in [0, z]$*.*
3. *If* $\delta < \underline{\delta}$*, the social planner chooses* $c_t^* = 0$*.*

An aggregate contribution above the threshold is wasteful, which means that the social planner chooses the aggregate contribution share c_t such that $c_t \leq z \; \forall t$. It follows from Equation 1 that $e_{t+1} = e_t - g\,(zne_t - c_t e_t) \; \forall t$. To find the social planner's solution, we can therefore maximize the value $V(e_t)$ depending on consumption in t and the discounted future value $V(e_{t+1})$ in the following way:

$$\begin{aligned} \max_{c_t} V(e_t) = ne_t - c_t ne_t + \delta V(e_{t+1}) \\ \text{s.t.} \\ c_t \leq z \\ c_t \geq 0 \\ e_{t+1} = e_t - g(zne_t - c_t ne_t)\,. \end{aligned} \tag{A.1}$$

Maximizing the corresponding Lagrangian

$$\max_{c_t} V(e_t) = ne_t - c_t ne_t + \delta V(e_t - gzne_t + gc_t ne_t) - \lambda_{1,t}(c_t - z) + \lambda_{2,t} c_t,$$

we get a first order condition w.r.t. c_t (the FOC), an envelope condition capturing the derivative of the value function with respect to the endowment (the EC) and two complementary slackness conditions:

$$\begin{aligned} \text{(FOC)} \qquad -ne_t + \delta V'(e_{t+1}) gne_t - \lambda_{1,t} + \lambda_{2,t} &= 0 \\ \text{(EC)} \qquad V'(e_t) &= n - nc_t + \delta V'(e_{t+1})(1 - gnz + gnc_t) \\ \lambda_{1,t}(c_t - z) &= 0 \\ \lambda_{2,t} c_t &= 0 \\ c_t - z &\leq 0 \\ c_t &\geq 0 \\ \lambda_{1,t}, \lambda_{2,t} &\geq 0\,. \end{aligned} \tag{A.2}$$

We distinguish three cases. First, suppose that the condition $c_t \leq z$ is binding. In this case, $c_t = z$, and $\lambda_{2,t} = 0$. Moreover, for $c_t = z$ it holds that $e_{t+1} = e_t$. The FOC and EC in System of Equations A.2 become $\delta V'(e_t) gne_t = ne_t + \lambda_{1,t}$ and $V'(e_t) = n - nz + \delta V'(e_t)$. Solving for $\lambda_{1,t}$ and $V'(e_t)$ yields $V'(e_t) = \frac{n(1-z)}{1-\delta}$ and $\lambda_{1,t} = \frac{\delta g n^2 e_t (1-z)}{1-\delta} - ne_t$. The condition $\lambda_{1,t} \geq 0$ is then equivalent to $\delta \geq \frac{1}{1+ng(1-z)}$. Hence, $c_t = z$ is a solution iff $\delta \geq \frac{1}{1+ng(1-z)} (= \underline{\delta})$.

Second, suppose that the condition $c_t \geq 0$ is binding, and in consequence that $c_t = 0$, $\lambda_{1,t} = 0$. It follows for the FOC and the EC in System of Equations A.2 that $-ne_t + \delta V'(e_{t+1}) gne_t + \lambda_{2,t} = 0$ and $V'(e_t) = n + \delta V'(e_{t+1})(1 - gnz)$. We guess and verify that the value function $V(e_t)$ is linear in e_t: $V(e_t) = ke_t$, with k being a positive constant. This implies $V'(e_t) = V'(e_{t+1}) = k$. We can then use the EC to determine that $k = V'(e_t) = \frac{n}{1-\delta+\delta gnz}$. Using the FOC we determine that $\lambda_{2,t} = ne_t - \frac{\delta g n^2 e_t}{1-\delta+\delta gnz}$. The condition $\lambda_{2,t} \geq 0$ is then equivalent to $\delta \leq \frac{1}{1+ng(1-z)} (= \underline{\delta})$. Thus, $c_t = 0$ is a viable solution to our problem iff $\delta \leq \underline{\delta}$.

Third, consider an interior solution $c_t \in [0, z]$. This is the case if $\lambda_{1,t} = 0$ and $\lambda_{2,t} = 0$, which leads to the following FOC and EC: $\delta V'(e_{t+1}) g e_t n = e_t n$ and $V'(e_t) = n(1 - c_t) + \delta V'(e_{t+1})(1 -$

$gzn + gc_t n)$. The FOC implies that $V'(e_t) = \frac{1}{\delta g}$, $\forall t$, such that the EC holds independently of c_t. It follows for $\delta n g(1-z) + \delta = 1$, or equivalently for $\delta = \underline{\delta}$, that any contribution level above or equal to zero and below or equal to the sustainability threshold z is optimal. These three results together establish Proposition A.1.

A.2 Proof of Proposition 2: Markov Perfect Equilibria

We prove Proposition 2 here. For $n = 1$ the solution of the social planner's problem, Proposition A.1, applies. Observing that for $n = 1$, we have $\underline{\delta} = \bar{\delta} = \bar{\bar{\delta}}$, it follows that a single player chooses *Defection* for $\delta < \bar{\delta}$; and *Cooperation* for $\delta > \bar{\bar{\delta}}$, as claimed by Proposition 2. Suppose from now on that $n > 1$.

We first show that there is no MPE with waste. Assume the contrary. Then there is a time t at which the sum of contributions is above the threshold, i.e., $\sum_{i=1}^{n} c_{i,t} e_t > Z_t$. A single player l could deviate by slightly reducing her contribution to $c'_{l,t} = c_{l,t} - \varepsilon$, with $\varepsilon > 0$ and $\varepsilon e_t \leq \sum_{i=1}^{n} c_{i,t} e_t - Z_t$. Her benefits in time t increase, while the state variable e_{t+1} stays constant. As Markov strategies only depend on the state variable, the continuation of the game is unchanged. Hence, reducing her contribution is a strict improvement for player l, in contradiction to the assumption of the original situation being a MPE. Moreover, using the same argument, if there is a beneficial deviation that yields waste, there is a more attractive deviation that does not exceed the aggregate threshold.

Player i maximizes (2), taking the contributions of the others as given. Let $\bar{c}_t$ denote the average contribution share of the other players. Using that the overall contributions never exceed the threshold, player i maximizes

$$
\begin{aligned}
\max_{c_{i,t}} V(e_t) = e_t - c_{i,t} e_t + \delta V(e_{t+1}) \\
\text{s.t.} \\
c_{i,t} e_t + (n-1)\bar{c}_t e_t \leq zne_t \\
c_{i,t} \geq 0 \\
e_{t+1} = e_t - g(zne_t - c_{i,t} e_t - (n-1)\bar{c}_t e_t)\,.
\end{aligned}
\tag{A.3}
$$

Taking derivatives of the corresponding Lagrangian

$$
\begin{aligned}
\max_{c_{i,t}} V(e_t) = e_t - c_{i,t} e_t + \delta V(e_t - gzne_t + gc_{i,t} e_t + g(n-1)\bar{c}_t e_t) \\
-\lambda_{1,t} e_t (c_{i,t} + (n-1)\bar{c}_t - zn) + \lambda_{2,t} c_{i,t}
\end{aligned}
$$

leads to the following first order condition (FOC), envelope condition (EC) and two complementary slackness conditions:

$$
\begin{aligned}
\text{(FOC)} \qquad -e_t + \delta V'(e_{t+1}) g e_t - \lambda_{1,t} + \lambda_{2,t} &= 0 \\
\text{(EC)} \qquad V'(e_t) &= 1 - c_{i,t} + \delta V'(e_{t+1})(1 - gzn + gc_{i,t} + g\bar{c}_t(n-1)) \\
\lambda_{1,t} e_t (c_{i,t} + (n-1)\bar{c}_t - zn) &= 0 \\
\lambda_{2,t} c_{i,t} &= 0 \\
c_{i,t} + (n-1)\bar{c}_t - zn &\leq 0 \\
c_{i,t} &\geq 0 \\
\lambda_{1,t}, \lambda_{2,t} &\geq 0.
\end{aligned}
\tag{A.4}
$$

If the endowments are zero, $e_t = 0$, then contributions in all further periods are zero for any strategy profile. Hence, we focus on $e_t > 0$. Then the first complementary slackness condition simplifies to $\lambda_{1,t}(c_{i,t} + (n-1)\bar{c}_t - zn) = 0$. We search for symmetric equilibria, meaning that all players contribute the same fraction of their endowment

First, suppose that $c_{i,t} \geq 0$ is binding. In this case it holds that $c_{i,t} = 0$ and the first complementary slackness condition further simplifies to $\lambda_{1,t}((n-1)\bar{c}_t - zn) = 0$. This is satisfied either if $(n-1)\bar{c}_t = zn$ or $\lambda_{1,t} = 0$. The former subcase cannot constitute a symmetric equilibrium, as $c_{i,t} = 0$ while $\bar{c}_t = \frac{nz}{n-1} > 0$. Consider hence the latter subcase: $\lambda_{1,t} = 0$. It follows for the FOC and the EC that:

$$\begin{aligned} &\text{(FOC)} \qquad -e_t + \delta V'(e_{t+1}) g e_t + \lambda_{2,t} = 0 \\ &\text{(EC)} \qquad V'(e_t) = 1 + \delta V'(e_{t+1})\big(1 - gzn + g\bar{c}_t(n-1)\big). \end{aligned} \tag{A.5}$$

We guess and verify that the value function $V(e_t)$ is linear in e_t: $V(e_t) = ke_t$, with k being a positive constant. This implies $V'(e_t) = V'(e_{t+1}) = k$. We can then use the EC in System of Equations A.5 to determine k and the FOC to determine $\lambda_{2,t}$:

$$\begin{aligned} k = V'(e_t) &= \frac{1}{1-\delta+\delta gzn}, \\ \lambda_{2,t} &= \left(1 - \frac{\delta g}{1-\delta+\delta gzn}\right) e_t. \end{aligned} \tag{A.6}$$

With $V(e_t) = \frac{1}{1-\delta+\delta gzn} e_t$ we found a consistent solution for the value function that satisfies the FOC and the EC. The condition $\lambda_{2,t} \geq 0$ is then equivalent to $\delta g(1-zn) \leq 1-\delta$. Thus, $c_t = 0$ is a symmetric MPE iff $\delta \leq \frac{1}{1+g(1-zn)} (= \bar{\bar{\delta}})$.

Second, suppose that $c_{i,t} e_t \leq nze_t - \bar{c}_t(n-1)e_t$ is binding. In this case it holds that $c_{i,t} = nz - \bar{c}_t(n-1)$. The second complementary slackness condition, $\lambda_{2,t} c_{i,t} = 0$, is satisfied either if $c_{i,t} = 0$ or if $\lambda_{2,t} = 0$. The former subcase cannot constitute a symmetric equilibrium, as $c_{i,t} = 0$ while $\bar{c}_t(n-1) + c_{i,t} = nz$ implies $\bar{c}_t = \frac{nz}{n-1} > 0$. Consider hence the latter subcase: $\lambda_{2,t} = 0$. It follows for the FOC and the EC that

$$\begin{aligned} &\text{(FOC)} \qquad -e_t + \delta V'(e_{t+1}) g e_t - \lambda_{1,t} = 0 \\ &\text{(EC)} \qquad V'(e_t) = 1 - zn + \bar{c}_t(n-1) + \delta V'(e_{t+1}). \end{aligned} \tag{A.7}$$

For $c_{i,t} = nz - \bar{c}_t(n-1)$, endowments are constant, such that $e_{t+1} = e_t$. We can solve the System of Equations A.7 for $\lambda_{1,t}$ and $V'(e_t)$ using that it follows from symmetry and $c_{i,t} = nz - \bar{c}_t(n-1)$ that $c_{i,t} = \bar{c}_t = z$:

$$\begin{aligned} V'(e_t) &= \frac{1-z}{1-\delta} \\ \lambda_{1,t} &= \left(\frac{\delta g(1-z)}{1-\delta} - 1\right) e_t. \end{aligned} \tag{A.8}$$

The condition $\lambda_{1,t} \geq 0$ is then equivalent to $\delta \geq \frac{1}{1+g(1-z)} (= \bar{\delta})$. Thus, $c_t = z$ is a symmetric MPE iff $\delta \geq \bar{\delta}$.

Third, suppose that we have an interior solution, in which case it holds that $\lambda_{1,t} = 0$ and $\lambda_{2,t} = 0$. The FOC and EC from System of Equations A.5 become

$$\begin{aligned} &\text{(FOC)} \qquad \delta V'(e_{t+1}) g e_t = e_t, \\ &\text{(EC)} \qquad V'(e_t) = 1 - c_{i,t} + \delta V'(e_{t+1})\big(1 - gzn + gc_{i,t} + g\bar{c}_t(n-1)\big). \end{aligned} \tag{A.9}$$

It follows from the FOC that $V'(e_{t+1}) = \frac{1}{\delta g} \forall t$. Plugging this result into the EC, while using symmetry in the sense that $c_{i,t} = \bar{c}_t = \tilde{c}$, we find:

$$\tilde{c} = \frac{1-\delta+\delta g(zn-1)}{\delta g(n-1)}. \tag{A.10}$$

Note that this is a consistent solution only if $0 \leq \tilde{c} \leq z$, which is the case whenever $\delta \geq \frac{1}{1+g(1-z)} (= \bar{\delta})$ and $\delta \leq \frac{1}{1+g(1-zn)} (= \bar{\bar{\delta}})$ hold. Observe that this *Inbetween* strategy coincides with *Defection* for $\delta = \bar{\delta}$ and with *Cooperation* for $\delta = \bar{\bar{\delta}}$; while it yields an intermediate value $\tilde{c} \in (0, z)$ for $\delta \in (\bar{\delta}, \bar{\bar{\delta}})$.

These three results together establish Proposition 2.

B Appendix: Additional Figures and Tables

B.1 Evolution of endowments

Table B.1: Treatment effects and endowment time trend

Dependent Variable:	Endowment	
Model:	(1)	(2)
Variables		
Period	-11.39***	-11.39***
	(1.144)	(1.146)
T-OnePlayer × Period	3.883**	3.883**
	(1.616)	(1.618)
T-LowThreshold × Period	6.139***	6.139***
	(1.311)	(1.313)
T-OnePlayer	-2.129	0.2379
	(1.443)	(2.505)
T-LowThreshold	-3.076***	-2.020
	(1.117)	(1.808)
Constant	107.7***	135.3***
	(1.007)	(17.71)
Controls	No	Yes
Fit statistics		
Observations	2,790	2,790
R^2	0.24353	0.28172
Wald (joint nullity)	42.612	21.592

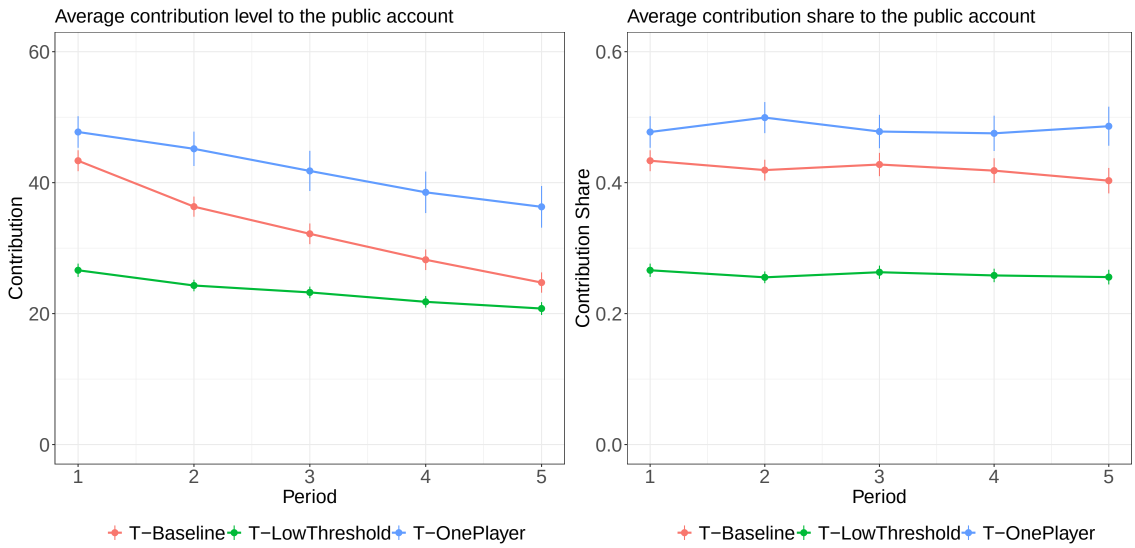
Clustered (Group) standard errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

B.2 Evolution of contributions

Figure B.1 shows the levels and shares of contributions over time, separated by treatment.

Table B.2 estimates the effect of *T-OnePlayer* and *T-LowThreshold* on the contribution shares to the public account. The dependent variable is considered as continuous, which allows us to run OLS regression in all columns. Removing strategic interactions increases contribution shares by about 4pp in period one and 8pp in period five. The effect is significant and amounts to about 6pp on average for all periods. The result is consistent with Hypothesis H-CoopDef-One: if more players choose *Cooperation* and fewer choose *Defection*, we can expect an increase in average contribution shares. The results also indicate that lowering the threshold reduces contribution shares by about 16pp. When comparing contribution shares in *T-LowThreshold* and *T-Baseline* there are two opposing effects at play. On the one hand, *T-LowThreshold* should increase average contribution shares, since fewer players choose *Defection*. On the other hand, we can expect a decrease of contribution shares since the socially efficient share is lower. Both effects seem to play a role, as we observe a decline in average contribution shares of 16pp that is smaller than the reduction in the efficient contribution share of 25pp.

Figure B.1: Contribution to the public account (levels and shares)



Notes: Mean and standard 95% confidence intervals, pooling groups and repetitions of the game

Table B.2: Treatment effects on Contribution Shares (OLS)

Dependent Variable:	Contribution Share					
	period 1		period 5		average 1-5	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
T-OnePlayer	0.0439*	0.0448*	0.0832***	0.0752***	0.0629***	0.0600***
	(0.0258)	(0.0254)	(0.0278)	(0.0274)	(0.0224)	(0.0223)
T-LowThreshold	-0.1673***	-0.1663***	-0.1473***	-0.1523***	-0.1606***	-0.1631***
	(0.0165)	(0.0165)	(0.0175)	(0.0165)	(0.0146)	(0.0139)
Constant	0.4335***	0.4500***	0.4030***	0.2891***	0.4203***	0.3464***
	(0.0153)	(0.0793)	(0.0156)	(0.0703)	(0.0137)	(0.0600)
Controls	No	Yes	No	Yes	No	Yes
Fit statistics						
Observations	1,692	1,692	1,692	1,692	1,692	1,692
R^2	0.18750	0.19261	0.13113	0.16136	0.28520	0.30834
Wald (joint nullity)	89.560	28.954	70.348	24.533	122.59	38.345

Clustered (Group) standard errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

B.3 Cognitive Ability and Wasteful Behavior

Table B.3: Effect of Raven on Overcontribution dummy

Dependent Variable:	Overcontribution dummy					
	period 1		period 5		average 1-5	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
	Logit	Logit	Logit	Logit	OLS	OLS
Variables						
Raven	-0.1579***	-0.0870	-0.1050***	-0.0761**	-0.0222***	-0.0156**
	(0.0314)	(0.0545)	(0.0240)	(0.0362)	(0.0043)	(0.0063)
T-OnePlayer × Raven		-0.0904		0.0095		-0.0031
		(0.1147)		(0.0978)		(0.0126)
T-LowThreshold × Raven		-0.1194*		-0.0560		-0.0127
		(0.0687)		(0.0505)		(0.0093)
Constant	-1.270	-1.779**	-2.354***	-2.623***	0.1243	0.0654
	(0.8301)	(0.8641)	(0.6772)	(0.7219)	(0.1300)	(0.1373)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	1,692	1,692	1,692	1,692	1,692	1,692
R^2					0.10152	0.10464
Pseudo R^2	0.06042	0.06506	0.05482	0.05608		
Wald (joint nullity)	4.5983	5.0249	7.5497	6.1891	8.6095	7.4970

Clustered (Group) standard errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Overcontribution is defined as 1 if the participant's contribution share $c_{i,t}$ is strictly larger than the sustainability threshold parameter z and 0 otherwise.

Table B.4: Effect of Raven on Waste dummy

Dependent Variable:	Waste dummy					
	period 1		period 5		average 1-5	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
	Logit	Logit	Logit	Logit	OLS	OLS
Variables						
Raven	-0.1308*	-0.1656	-0.1286*	-0.2310**	-0.0251***	-0.0363*
	(0.0737)	(0.1224)	(0.0715)	(0.1083)	(0.0094)	(0.0185)
T-OnePlayer × Raven		-0.0669		0.1338		0.0131
		(0.1662)		(0.1441)		(0.0205)
T-LowThreshold × Raven		0.2186		0.1293		0.0186
		(0.1825)		(0.1831)		(0.0331)
Constant	-1.633	-1.385	-1.057	-0.3703	0.3355	0.4138
	(1.784)	(1.930)	(1.829)	(1.884)	(0.2736)	(0.2962)
Treatmentdummies	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	558	558	558	558	558	558
R^2					0.18387	0.18565
Pseudo R^2	0.10916	0.11786	0.06902	0.07094		
Wald (joint nullity)	3.7497	3.0584	3.2101	2.9939	6.7304	6.1179

Clustered (Group) standard errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Waste is defined as 1 if the group's aggregate contribution $\sum_{i=1}^{n} c_{i,t}e_t$ is strictly larger than the sustainability threshold Z_t and 0 otherwise.

B.4 Association with New Ecological Paradigm (NEP)

Table B.5: Effect of NEP on playing Close to Cooperation

Dependent Variable:	Close to Cooperation					
	period 1		period 5		average 1-5	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
	Logit	Logit	Logit	Logit	OLS	OLS
Variables						
NEP	0.0279*	-0.0059	0.0283**	0.0038	0.0046**	-0.0002
	(0.0147)	(0.0225)	(0.0123)	(0.0256)	(0.0022)	(0.0037)
T-OnePlayer × NEP		0.0973**		0.0419		0.0099*
		(0.0379)		(0.0400)		(0.0057)
T-LowThreshold × NEP		0.0427		0.0326		0.0065
		(0.0295)		(0.0283)		(0.0046)
Constant	-3.031**	-1.462	-2.806***	-1.503	-0.0770	0.1617
	(1.327)	(1.506)	(1.024)	(1.473)	(0.1914)	(0.2180)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	1,692	1,692	1,692	1,692	1,692	1,692
R^2					0.14517	0.15137
Pseudo R^2	0.09881	0.10619	0.07978	0.08198		
Wald (joint nullity)	5.2641	4.9003	4.8565	4.7357	7.3428	7.5625

Clustered (Group) standard errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: NEP measures pro-environmental orientation (Dunlap et al., 2000).

Table B.6: Effect of NEP on playing Close to Defection

Dependent Variable:	Close to Defection					
	period 1		period 5		average 1-5	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
	Logit	Logit	Logit	Logit	OLS	OLS
Variables						
NEP	-0.0441**	-0.0163	-0.0271**	-0.0329*	-0.0018**	-0.0020
	(0.0214)	(0.0327)	(0.0128)	(0.0180)	(0.0008)	(0.0018)
T-OnePlayer × NEP		-0.0055		0.0558		0.0006
		(0.0482)		(0.0603)		(0.0021)
T-LowThreshold × NEP		-0.0581		0.0040		0.0002
		(0.0396)		(0.0250)		(0.0022)
Constant	5.252**	3.807	3.406**	3.587**	0.4819***	0.4887***
	(2.303)	(2.403)	(1.447)	(1.707)	(0.1261)	(0.1522)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	1,692	1,692	1,692	1,692	1,692	1,692
R^2					0.07813	0.07819
Pseudo R^2	0.09485	0.09993	0.08722	0.08848		
Wald (joint nullity)	3.4517	3.0074	5.4242	4.9883	3.2490	2.8520

Clustered (Group) standard errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

References

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2017): "When should you adjust standard errors for clustering?" Tech. rep., National Bureau of Economic Research.

Battaglini, M., S. Nunnari, and T. R. Palfrey (2016): "The Dynamic Free Rider Problem: A Laboratory Study," *American Economic Journal: Microeconomics*, 8, 268–308.

Berger, S. and A. M. Wyss (2021): "Measuring pro-environmental behavior using the carbon emission task," *Journal of Environmental Psychology*, 75, 101613.

Cadsby, C. B. and E. Maynes (1999): "Voluntary provision of threshold public goods with continuous contributions: experimental evidence," *Journal of Public Economics*, 71, 53–73.

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

Dal Bó, P. and G. R. Fréchette (2018): "On the determinants of cooperation in infinitely repeated games: A survey," *Journal of Economic Literature*, 56, 60–114.

Dunlap, R., K. V. Liere, A. Mertig, and R. E. Jones (2000): "Measuring endorsement of the new ecological paradigm: A revised NEP scale," *Journal of Social Issues*, 56, 425–442.

Fréchette, G. R. and S. Yuksel (2017): "Infinitely repeated games in the laboratory: Four perspectives on discounting and random termination," *Experimental Economics*, 20, 279–308.

Fudenberg, D. and J. Tirole (1991): *Game Theory*, MIT press.

Ledyard, J. O. (1995): "Public Goods: A Survey of Experimental Research," in *Handbook of Experimental Economics*, Princeton University Press, 111–94.

Levhari, D. and L. J. Mirman (1980): "The great fish war: an example using a dynamic Cournot-Nash solution," *The Bell Journal of Economics*, 322–334.

Maples-Keller, J. L., R. L. Williamson, C. E. Sleep, N. T. Carter, W. K. Campbell, and J. D. Miller (2019): "Using Item Response Theory to Develop a 60-Item Representation of the NEO PI–R Using the International Personality Item Pool: Development of the IPIP–NEO–60," *Journal of Personality Assessment*, 101, 4–15, pMID: 29087223.

Ostrom, E. (1990): *Governing the commons: The evolution of institutions for collective action*, Cambridge university press.

Portner, H. O., D. C. Roberts, H. Adams, C. Adler, P. Aldunce, E. Ali, R. A. Begum, R. Betts, R. B. Kerr, R. Biesbroek, et al. (2022): "Climate change 2022: impacts, adaptation and vulnerability," .

Proto, E., A. Rustichini, and A. Sofianos (2019): "Intelligence, personality, and gains from cooperation in repeated interactions," *Journal of Political Economy*, 127, 1351–1390.

Vespa, E. (2020): "An experimental investigation of cooperation in the dynamic common pool game," *International Economic Review*, 61, 417–440.